Control and disposal of demonstratives, with electrophysiological evidence from English and Japanese

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

James M. Stevens

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jeanette K. Gundel, Yang Zhang

June, 2014

Acknowledgements

Foremost among my acknowledgements are to my advisors, Professors Jeanette K. Gundel and Yang Zhang. Dr. Gundel provided the willingness to engage my curiosity, however unformed, and the guidance through original scholarship and professional development. Thank you for the never-ending conversations, Jeanette. To Dr. Zhang I owe the prize of completing the research that lies at the heart of the dissertation, as without his very kind patience, expertise, and resources, most of what follows would have remained sketches in a notebook. Thank you, Yang, for being a mentor in the finest sense. Thank you, too, Professors Sera and Kac, who have both been a consistent and encouraging presence as teachers and guides. I would also like to acknowledge the invaluable assistance of Tess Koerner, Sharon Miller, and Edward Carney. I would also like to recognize various funding sources that allowed for the present work to happen, including University of Minnesota start-up funds to Professor Zhang and funding from the University of Minnesota Linguistics Program, College of Liberal Arts, the FLAS Scholarship, and the Instructional Technology Fellowship.

Dedication

This thesis is dedicated to my partner in this life, Candance, as very little of what I have done or become would be so without her support, love, patience, and this and that.

A signal is a means by which one animal makes use of another animal's muscle power.
-Richard Dawkins, *The Selfish Gene*

Whatever exists has already been named,
and what humanity is has been known;
no one can contend
with someone who is stronger.
The more the words,
the less the meaning,
and how does that profit anyone?
-*Ecclesiastes 6:10-11*

Abstract

Demonstratives are lexical forms that pick out an object by making use of constraints in a discourse context to establish some form of contrast. They represent among the most basic uses of language, yet escape simple definition. The two forms “this” and “that” in English, the three forms こ “ko,” そ “so,” あ “a” in Japanese, etc. have traditionally been interpreted in terms of relative distance from the speaker or hearer to the object. An alternative framework presents demonstratives as a device to refocus attention (e.g., Strauss, 2002), where “this” requires more of the hearer’s attention, to represent new information, or to refer to objects relatively important in the discourse, versus objects called “that” or “it”. However, distance or attention alone is an insufficient parameter to predict form selection. This dissertation builds upon Brovold and Grush’s (2012) analysis of demonstratives in terms of control over an object. I propose that demonstrative forms, at least in English and Japanese, can be predicted from an array of *control spaces* that allow for varying levels of potential action toward an object. The proposed framework of control spaces implies an embodied view of language, such that language and physical behavior work toward shared goals and link mature human language use to other forms of animal communication, to child language acquisition, and with adult humans’ response to a changing environment (De Ruiter, 2006; Melinger & Levelt, 2005).

The control space framework incorporates previous models and makes three specific predictions. First, relative distance serves as one determinant for demonstrative form selection. Second, the weight of relative distance in demonstrative form selection

will decrease without shared eye gaze between the speaker and hearer. 3. Demonstrative use would show as a tendency greater dependency on the co-speech pointing gesture for speakers of a language (e.g., English) that has fewer demonstrative terms than another (e.g., Japanese). To evaluate the three predictions, two perceptual studies that employed behavioral and event-related potential (ERP) measures were conducted to assess language users' intuitions and cortical-level neural responses to the presentation of demonstrative expressions in differing visual contexts. When the simultaneous visual context with demonstrative expressions did not correspond with expected relative distances among speaker, hearer, and object, participants responded with significantly longer reaction times accompanied by a brain response called the *N400*, which is associated with semantic/contextual incongruities (Kutas & Hillyard, 1980, 1984). The elicitation of the N400 response for the incongruent audiovisual matchup in demonstrative use depended on shared gaze between the speaker and hearer, and the N400 responses were found in both English and Japanese subjects. English and Japanese subjects, however, differed in their responses to trials that did not include a pointing gesture in the visual scenes. In the absence of co-speech pointing gesture, English speakers expressed a P600 response, an index of pattern violations, but Japanese speakers did not. These findings indicate that in addition to spatial distance, language users rely on shared gaze in determining the proper use of demonstrative forms and show language-specific sensitivity to the presence or absence of gesture when analyzing demonstratives.

The results of this dissertation project highlight the contingent, contrastive, and attention-orienting nature of demonstratives and further illustrate the necessity to study

speech communication as a multimodal social event, which is subject to a number of factors, including perspectives of the speaker and listener, physical context, gesture, and language per se. Demonstratives function as basic lexical means to make meaning out of the world. They represent a special case of names, which are interpreted here as deictic means to form an object out of a nameless ground. Demonstratives, more specifically, serve as names that are used once and then no longer refer to the same object, resuming the naming practice with every use. The ERP results reported in this dissertation show promising brain signature markers for understanding the multimodal processing nature of spatial demonstratives. Further directions for research are suggested, including measuring the relationship between autistic individuals' competence with demonstratives alongside their imitative physical skills.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Demonstratives in space

1.1. Introduction.

The study of language holds particular fascination in its apparent unique place among humans and invites further questions about the nature of cognition itself. Several authors have discussed the role of language among a set of tools to simplify a search space (Bowerman, 1996; Clark, 1997; Demuynck, Duchateau, Van Compernolle, & Wambacq, 2000; Hollich et al., 2000; Sperber & Wilson, 2005). Two broad and complementary means of constraining a search space are description and demonstration. Language can identify a referent by description, which is essentially a reconstruction of previous demonstrations. "The first U.S. President born in Hawaii" helps narrow a search in our world to a single entity by description without additional cues. Conversely, the expression "the senator over there" is in itself insufficient to identify a referent. In order to understand which senator is of interest, we expect in addition to this under-specified description such additional physical cues as a pointing gesture, the direction of the speaker's gaze, and the relative distance between our conversation and the senators in the vicinity.

The physical constraints of demonstrative use have drawn considerable attention in fields as diverse as philosophy, linguistics, cognitive science, anthropology, and child psychology. In what follows, I will discuss demonstrative use in physiological and ritualistic terms. The physiological account will seek to describe what we do when demonstratives are used. The ritualistic account will aim to describe what demonstratives do when they are used; this latter discussion will be saved until the final chapter. In that last chapter, I will develop a definition of demonstratives as single-use names, meaning that demonstratives have a naming function, by which one can identify an object with a symbol and afterward dispose with the name itself. The definition of demonstratives is difficult, as will be discussed in 1.2.

The bulk of this dissertation will describe demonstrative use in perceived space (exophoric usage). I will argue that a demonstration is an action-oriented behavior that expresses possibilities to interact with a referent. Specifically, I will posit that demonstrative use establishes a ***control space*** (cf. Brovold & Grush, 2012; Langacker, 2002, 2009), space framed by physical and linguistic means that reflect the imagined action possibilities afforded to the interlocutors (cf. Gibson, 1977; Norman, 1988). A control space can be interpreted as having physical dimensions for material interaction, e.g., grasping a pencil, but can also be a non-material space formed through the use of pitch, syntax, memory, etc. I will also argue that control spaces can be ordered according to the level of control presumed possible by the speaker, and that the ordered control spaces are encoded by expected terms, e.g., in English, “this” for the first control space, “that” for the second. If, for instance, the speaker and hearer do not have the same vantage point and thus different capacities to interact with the environment, a space in which the speaker realizes more immediate possibilities for interaction with an object would be the *first* control space; a space where the hearer would have more immediate possibilities to interact with a referent would be the *second* control space. If the speaker and hearer have a shared point of view, then the possibilities for interacting with the object would follow other considerations, such as the distance from both interlocutors.

Three hypotheses will be developed out of the use of control space to describe spatial demonstrative use.

Hypothesis 1: The properties of an individual language, such as the number of its demonstrative terms, will influence the expectation of physical constraints on a demonstration, such that language groups will differ in their responses to the presence or absence of a co-speech pointing gesture, among other forms of extra-linguistic communication. Although humans universally have an interest in assessing means to control their environment, how they categorize the available action possibilities need not be universal, given the peculiarities of each language and the language group's expected use of extra-linguistic tools. This hypothesis serves as an extrapolation from the Mutually Adaptive Modalities Hypothesis (De Ruiter, 2006; Melinger & Levelt, 2005), which predicts that language and gesture use will compensate for each other to maximize communication. The Mutually Adaptive Modalities hypothesis will be discussed in Chapter 2, but briefly put this hypothesis posits that more detailed verbal information lead to a decrease in gesture and vice versa. Although the hypothesis was formulated in terms of use within a given language, e.g., in the case of aphasia patients, a trend toward less reliance on extra-linguistic tools such as gesture in proportion to the richness of the demonstrative system may be observed across languages, although individual language groups (e.g., Italian) may be exceptional in the expectation for gesture. Hearers therefore can negotiate a search space with fewer physical cues to find a referent if the lexical information is sufficiently rich.

Hypothesis 2: Demonstrative use is sensitive to the relative distance of the referent to the speaker and hearer, since interlocutors will be able to interact differently with the referent based on the spatial context, and the role of relative distance as a determinant of demonstrative form.

Relative distance from the speaker to the referent object itself is a factor to which demonstrative form selection is sensitive, according to the Control Space Model, because the speaker will have more possibilities by which to control an object if that object is physically near the speaker, other variables being equal (e.g., one can otherwise imagine a near object hidden behind something else and thus less available for interaction). When the object is nearest the speaker, we will hypothesize a *first* control space that allows for a maximum amount of action possibilities for the speaker relative to other agents or distracter objects and therefore expect a correlated demonstrative form.

A key part of this hypothesis is that in English the so-called proximal demonstrative form will correlate with an object that is near the speaker in relative terms. That is, an object might be found one meter from the speaker, a distance that might be considered small in terms of everyday human activities. However, if the object is sitting on the hearer's lap, even an otherwise short distance such as one meter would be relatively longer compared to the distance of the object to the hearer. When in situation similar to a referent object lying on top of the hearer's lap, the speaker would estimate having fewer action possibilities to interact with the object relatively to the hearer. In English, the expected demonstrative form in the latter context would be "that," and in Japanese the expected form will be "so," as the form correlates with the *second* control space.

Not only could the delineation of a control space be determined relatively to the speaker and hearer but also compared to other objects. We can refer to a marsupial as "this marsupial," to contrast to other pouched objects that are not in our focus and thus have fewer action possibilities associated with them.

Hypothesis 3: Relative distance would exert less importance in the selection of a demonstrative form if the speaker and hearer do not have shared gaze, because the speaker's ability to interact with the referent will be increased relatively to the hearer if the hearer is not attending in the same direction..

It has been mentioned that relative distance alone cannot be the sole determining factor for demonstrative from selection. Hypothesis 3 focuses on one situation that differs from the use of distance to decide a demonstrative, as distance will be less important if the hearer is not aware of the referent object's location. This hypothesis seeks to include situations where the hearer's attention is redirected. For example, a speaker might refer to a book sitting next to the hearer, whose eyes are fixated on a TV screen, by saying, "Oh, what's this book here?" The proximal demonstrative "this" may be considered acceptable despite the referent being nearer the hearer, as the speaker may consider more action possibilities toward the object than the hearer, who does not appear inclined to interact with the object spontaneously. The referent object therefore can still be considered within a first control space.

I will present two event-related potential (ERP) studies that use an audio-visual paradigm to test English or Japanese-speaking participants' expectations for the pairing of a demonstrative determiner (e.g., “that flamingo”) with the simultaneous presentation of a picture that includes the images of a speaker, a hearer, and a referent. The ERPs are a time-locked measure of voltage changes on the scalp in response to compared stimuli, and the present project analyzed these voltage changes by varying such parameters as demonstrative forms, relative distances between the speaker and hearer from the referent, the use of a pointing gesture, and the use of shared gaze. The demonstrative forms presented in English were “this” and “that” and in Japanese were こ(“ko”), そ(“so”), and あ(“a”). One main finding was that in the English-speaking subject group an ERP response called the *P600* was elicited when the pairing of a demonstrative form was unexpected by participants for a visual scene that did not include a pointing gesture, whereas no such effect was observed from the Japanese subject group for identical visual scenes. The P600 has been described as an index for pattern incongruities (Hagoort, Brown, & Groothusen, 1993; Hagoort, Brown, & Osterhout, 1999; Osterhout & Holcomb, 1992; Schlesewsky & Bornkessel, 2006; Van Herten, Chwilla, & Kolk, 2006). Another finding in both subject groups is that both reaction time data and ERP results showed participants’ sensitivity to relative distance as a factor in the expectation of a demonstrative form. A different ERP, the *N400*, was elicited by an unexpected pairing of a demonstrative form with a visual scene when the relative distance was varied. The N400 has been observed to indicate semantic violations in context (Brown & Hagoort, 1993; Kutas & Federmeier, 2011; Kutas & Hillyard, 1980, 1984). Participants in the present study in both subject groups showed sensitivity to the referent’s distance from both the speaker and the hearer. Notably, the N400 effect was not elicited when the speaker and hearer did not have shared gaze, irrespective of their relative distances to the referent. More details on the ERP P600 and N400 responses, including literature review, method, results in the present investigation, and discussion, are provided in Chapters 3 and 4. Modified versions of these two chapters have been peer reviewed and accepted for publication in the Journal of Neurolinguistics (Stevens & Zhang, 2013; Stevens & Zhang, 2014).

1.2. Demonstratives and the near and far problem.

Diessel (1999) provided three criteria for demonstratives. Firstly, demonstratives are *deictic* expressions with specific syntactic roles. *Deictic* terms, similarly to demonstratives, escape simple definition, but for the time being we can consider them terms that require a point of reference to be meaningful. The words “I,” “yesterday,” “go,” and “here” all require reference to some aspect of the discourse context to be understood. The deictic features of demonstratives, according to Diessel’s cross-linguistic analysis, include the distance of the referent to a deictic center (that is, some point of reference, such as the speaker), whether or not the referent was visible, if it was uphill or downhill, and whether moving toward or away from the deictic center. The syntactic roles performed by demonstratives include that of independent pronouns (e.g., “this”), noun determiners (e.g., “this gym”), adverbs (e.g., “here”), and in some languages a non-verbal copular use (e.g., in French, “Voici votre repas” = “Here’s your meal”). Demonstratives serve the pragmatic function of focusing another individual’s attention on some aspect of the speech situation, as well as to organize discourse information. The most basic function of demonstratives, according to Diessel, is “to orient the hearer outside of discourse in the surrounding situation” (p. 2). The last criterion is that demonstratives be characterized by specific semantic features, such that a deictic contrast is established between the forms within a language. Remarkably, Diessel's survey of 85 diverse languages found that all languages have a deictic contrast in their demonstrative forms, referring to entities that are near the deictic center versus some distance away from the center.

Demonstratives are often described in terms of spatial distance from the speaker (Anderson & Keenan, 1985; Bain, 1879; Greenberg, 1985; Lyons, 1977; Wu, 2004). Lyons (1977) recognizes the egocentricity of deictic speech in English by describing the interpretation of language in relation to a *canonical situation-of-utterance*, which places the speaker at the zero point of spatial and temporal dimensions. Thrane (1980) formalizes deictic expression with the use of features [±proximal] to differentiate "this" versus "that", [±space], and [±time] to describe demonstratives separately from the definite article. The notion of a distal form has been contested, as it suggests farther distance, yet what we call distal forms may be better considered associated with a hearer or in neutral space (Byron & Stoia, 2005; Enfield, 2003; Halliday & Hasan, 1979). Lyons takes "that" to be the unmarked form of the [±proximal] feature, rendering it more precisely a non-proximal form, which refers to that which is not close to the speaker but not necessarily far either. In fact, the difference between proximal and distal space might be considered neutralized, such that either form is possible, when a pointing gesture is present (Fillmore, 1971). These misgivings notwithstanding, the proximal-distal distinction does explain an intuition illustrated by the oddness of the demonstrative "that" used in (1).

(1) [*While tugging at the shirt one is wearing*] What do you think of this/??that shirt I found in my parents' basement?

Per the relative acceptability of (1), one could posit that "this" does denote objects that are closer to the speaker, especially if in direct contact. Distal demonstratives, on the other hand, seem more appropriate for objects close to a hearer, e.g., (2), where the proximal demonstrative "this" produces a less natural reading (unless perhaps the speaker is grasping the hearer's shirt) and could be judged to convey a derisive opinion.

(2) Is this/that the shirt you're wearing to the interview?

The deictic distinction in (1) and (2), according to Diessel's excellent survey of demonstrative systems, is far from unique. Indeed, a binary demonstrative system is the most frequently found among the world's language. One can imagine, given the recurrence of this system across languages, an argument for a biological foundation for spatial distinctions in language (Langacker, 1987, 2009; Regier, 1996).

An anatomical basis for a proximal-distal distinction could be forwarded on the basis of the human visual system (see Kemmerer, 1999, for an excellent review). An argument for anatomical categories of near and far space could be supported with evidence from brain-damaged patients showing spatial neglect, a neurological deficit whereby parts of the visual field cannot be consciously perceived (Cowey, Small, & Ellis, 1999; Halligan & Marshall, 1991; Heilman, Watson, Valenstein, & Damasio, 1983). Some patients have been observed to have a form of proximal-distal neglect, whereby near space can be consciously represented but the representation of far space is impaired or, conversely, far space can be represented with deficits in perceiving near space (Cowey, Small, & Ellis, 1994; Cowey et al., 1999; Halligan & Marshall, 1991). Based on these clinical findings, it could be hypothesized that the human brain divides up space into near and far domains analogously to how demonstratives seem to be used across languages with proximal and distal demonstrative terms. Kemmerer (1999) points out the studies of spatial neglect by Halligan and Marshall and Cowey, Small, and Ellis show considerable variability in patients' assessment of near and far space. Consequently, near and far visual domains likely intersect gradually, rather than forming a rigid boundary. To this author's knowledge, moreover, there have not been case reports of neglect patients with specific deficits in the use of demonstratives. One conclusion from studies of neglect that could guide our study of demonstratives is the simple observation that humans do divide up space into near and far domains (Farnè & Làdavas, 2002; Legrand, Brozzoli, Rossetti, & Farnè, 2007; Rizzolatti, Fogassi, & Gallese, 2002), a finding corroborated by non-human primate neuron stimulation studies (Iriki, Tanaka, & Iwamustongra, 1996) and experimental studies on healthy human subjects (Bjoertomt, Cowey, & Walsh, 2002; Gamberini, Seralgia, & Priftis, 2008).

There has been empirical evidence to support a perceptually-based proximal-distal division in language use (Coventry, Valdes, Castillo, & Guijarro-Fuentes, 2008; Stevens & Zhang, 2013). Coventry et al. found that the use of the proximal forms “this” in English and “este” in Spanish were highly correlated with the reach of one's hand or the reach with a hand tool versus longer distances. It was also found that the choice of demonstrative form was affected by who last manipulated the object, such that “this” was used more frequently if the experimental subject had been last to touch the object. The finding that the speaker last touching the referent can influence the choice of referential form is, interestingly, discussed as a means of ‘activating’ the object in the Givenness Hierarchy (Gundel, Hedberg, & Zacharski, 1993).

Coventry et al. argue for a perceptual basis for the binary proximal-distal system but leave contrastive uses in extrapersonal space (e.g., “this galaxy” and “that galaxy”) left unexplained. A second problem with the authors’ conclusion is that the use of a hand tool permitted the use of the proximal forms “this” and “este” for more distant spaces than without a tool, a finding that argues against a perceptual basis for the proximal-distal distinction. The influence, furthermore, of who had been last to touch the referent also does not coincide well with a distance model for demonstrative form selection. However, the findings in the Coventry et al. paper would be consistent with the argument of control space, such that physical distance is a factor for form selection, but the speaker’s use of a pointing tool extends the domain of control. The observation that if the speaker was last to touch the referent the proximal form is selected supports the role of control in the choice of demonstrative form.

Despite speakers’ intuition and any empirical evidence favoring a proximal-distal distinction, a description of demonstrative use primarily in terms of distance, e.g., [+proximal] as the denotation of “this” (cf. Lyons, 1977; Thrane, 1980), is overly simplistic. At least four reasons preclude a direct application of [±proximal] as the distinguishing feature between demonstrative forms. (*i*) Interlocutors typically make use of tools, including their bodies, which would be redundant if demonstrative forms denoted a [±proximal] value but in fact seem necessary for a felicitous demonstration. (*ii*)

Tool usage can facilitate a demonstrative form not licensed by spatial distance alone. (*iii*) One problem with using the [±proximal] feature is its vagueness. (*iv*) When demonstrative use does not correspond with physical distance, one can often evoke a metaphorical use to rescue some sense of distance, leading to the problem of ever knowing what is a literal and what is a metaphorical interpretation. In addition to the above concerns is the mentioned criticism that demonstratives can be used contrastively in extra-personal space, e.g., "this galaxy" (Kemmerer, 1999).

(*i*) Spatial proximity does not typically serve as the sole predictor of demonstrative form but more frequently is accompanied by some extra-linguistic cues. There are situations where the referent is salient enough that no physical indication is necessary, e.g., "This planet is very hospitable to sad lives" can be stated without pointing to the planet underfoot. However, the absence of any physical indication during a demonstration can often be odd. If a feature such as [+proximal] could serve as the denotation of "this," then a person with an ugly cat nestled behind her feet could look her listener in the eye, hands in her pockets, and refer felicitously to the animal snuggled against her heels as "this ugly cat." The cat is undoubtedly physically proximal, even touching the speaker, but the demonstration seems odd without the use of a gesture or a gaze in the direction of the referent. The direction of the speaker's gaze and the use of a pointing gesture would be redundant, if "this" denoted a proximal object.

(*ii*) The use of one's body or other tools can accompany a demonstrative form that would not be predicted by spatial proximity alone, as found in Coventry et al's (2008) findings. Frequently, a pointing gesture or other behavior can serve to redirect the attention of one's audience. When the speaker can appreciate a referent that is not apparently in the hearer’s awareness, a proximal form can be felicitous irrespective of spatial distance. As an example, imagine two young sisters, Alice and Angie, standing next to a bed getting ready for a wonderful evening. Alice tells Angie to pick out an outfit. One outfit is hanging on the door some distance away, and another is lying next to them on the bed. As Angie picks up the clothes lying on the bed, Alice expresses (3).

(3) [*Angie picks up the outfit on the bed*.]

Alice: No, not <u>that</u> one in your hands! You’d really do better with <u>this</u> one hanging on the door [*turning her head and waving an open palm toward the outfit hanging on the door*].

In (3), a distant object is picked out with the use of “this” when the hearer was not looking in the same direction. Notably, the demonstration in (3) would likely fail without the extra-linguistic cue of a waving hand, nod, or a pointing finger.

(*iii*) A concern with the use of a [±proximal] feature as the denotation of demonstrative forms is the vagueness of this feature. One common boundary between proximal and non-proximal space that has been discussed is that of one’s arm’s length, analogously to how neglect patients have discriminated near and far space (Kemmerer, 1999). However, no such demarcation in space can suit the variety of uses for the demonstrative forms. If one, for example, parks a car in front of a house with a “For Sale” sign in front, it would be unremarkable to ask, “What do you think of <u>this</u> house?,” even though the house is beyond one’s physical reach. In addition to the uses of a so-called proximal form for reference in extrapersonal space, there is an additional difficulty of knowing how to define a [±proximal] feature in languages with richer demonstrative systems (e.g., Japanese with three forms: "ko," "so," and "a").

(*iv*) The metaphorical use of demonstratives raises a concern for proximity to serve as the distinguishing feature of demonstrative forms, because it would require the hearer to switch between literal and metaphorical interpretations without systematic guidance. The above house example, where the speaker asked about "this house" seems, to this author's intuition, to suggest more openness to buying it than the more apparently neutral stance in asking, "What do you think of that house?" Such examples convey an attitude toward a referent in terms of what could be deemed emotional distance (Lakoff & Johnson, 1980; Lakoff, 1974; Semino & Culpeper, 2002; Wu, 2004). In (4), the use of the non-proximal form "that" is appropriate to communicate the speaker's emotional stance toward the referent.

(4) Yuck! What is that? [*shaking a bug off one's sleeve*]

One can similarly ask "What is that?" when holding up a foot to examine a stain on one's shoe. The use of "this" in either context would also be acceptable, but the availability of the distal form as a choice argues against distance as a unique determinant of demonstrative form selection. Interestingly, both contexts involve some displeasure on the part of the speaker, and this recognition would reasonably encourage separation from the source of displeasure. Although superficially the use of distance as the principal determinant of demonstrative forms seems to be rescued, the allowance of both literal and metaphorical interpretations of distance at random leads to the concern of never possibly being wrong. In (5), for instance, the proximal form "this" works despite the speaker being farther from the referent than the hearer.

(5) [*Son puts hand on car in garage*]

Father: [*standing behind son and farther from the car*]: Someday, <u>this</u> can be yours.

It is inviting to invoke a metaphorical sense in which the father in (5) is still somehow close to the car: because it is his property, because he is assuming the point of view of his son and is therefore vicariously close to the car, etc. This reasoning serves as a safety net for the distance model, such that any violation of the model can be construed as metaphorically close or distant from something. The ability to switch between literal and metaphorical interpretations *ad hoc* jeopardizes the predictive value of our explanation.

Given the above concerns with explaining demonstrative forms as denoting spatial distance from the speaker, other frameworks have been offered. In 1.3, we will examine the use of focus to differentiate English demonstrative use. In 1.4, we will discuss the role of control to explain the same phenomena.

1.3. Attentional models.

Another explanation for the different uses of demonstrative forms is the extent to which they demand attention. Peirce argued against treating demonstratives as indices because they are words and thus must maintain the same word meaning across uses and that this meaning for demonstratives is to "call attention to what is spoken of" (Fitzgerald, 1966:

59). Attentional models, framed in terms of "attention," "focus," or "deixis," have been proposed for Dutch (Kirsner & van Heuven, 1988; Kirsner, 1977), Afrikaans (Ponelis, 1993), Swahili (Leonard, 1995), and forwarded by Diessel (2006), having extensively surveyed typologically-diverse languages. The Givenness Hierarchy (Gundel, Hedberg, and Zacharski, 1993) similarly presents a framework that makes use of attentional status but differs from the above discussion in its treatment of referring forms generally, including definite and indefinite articles, as well as non-demonstrative pronominal forms. According to the Givennness Hierarchy, demonstrative pronouns "this" and "that" and pronominal determiner "this N" are correlated with the 'activated' cognitive status, whereby an object was recently brought to attention and might be considered in the speaker and hearer's short-term memory. The pronominal determiner "that N," on the other hand would require a less restrictive mental state, requiring only that the referent be presumed familiar to the hearer. The Givenness Hierarchy therefore should be noted for its explicit estimation of the hearer's cognitive state, including memory or attention, findings supported by corpus work in several languages and child language (Gundel & Johnson, 2013; Gundel, Ntelitheos, & Kowalsky, 2007). On the basis of extensive corpus work, Strauss (1993, 2002) describes demonstrative use in American English without the notion that "this" and "that" represent spatial terms. Strauss instead explains the difference between the English demonstrative forms in terms of *gradient focus*, and that demonstrative use can be better explained in terms of shared information between the speaker and hearer and the relative importance of a referent.

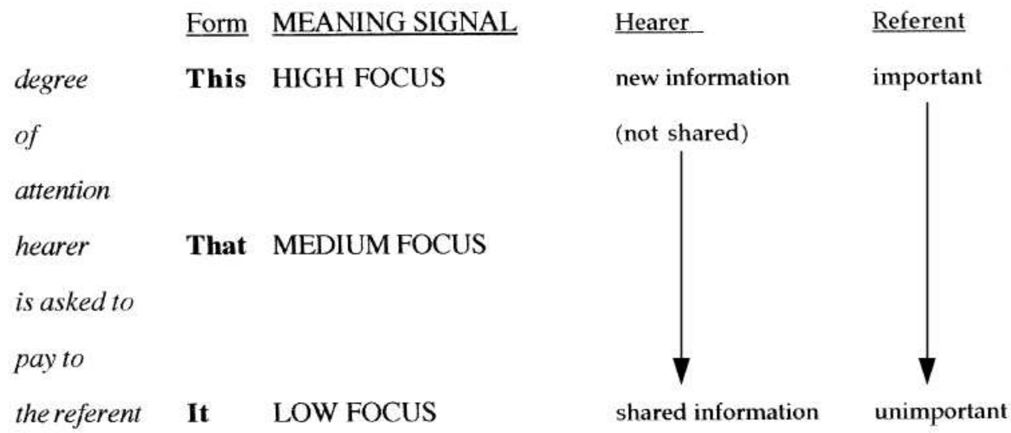

Fig. 1.1. The gradient focus model (Strauss, 2002).

In this model (Fig. 1.1), the demonstrative ‘this’ correlates with a HIGH FOCUS referent, ‘that’ with a MEDIUM FOCUS referent, and ‘it’ with a LOW FOCUS member, together forming a gradation of focus, which translates to ‘the degree of attention the hearer should pay to the referent’ (Strauss, 2002, p.135).

Strauss (2002) analyzed a 45,000 word corpus of spontaneously produced American English and found evidence against the traditional proximal-distal distinction and support for the model based on gradient focus. The number of tokens of the three target forms, ‘this,’ ‘that,’ and ‘it’ was 2076, but only 16% (333 tokens) of these were ‘this.’ Strauss argues against the proximal-distal model based on the alleged oddness of referring so seldom to objects that are close to the speaker. This argument, however, ignores the possibility that ‘that’ is used to refer to objects in non-proximal space, that is, ones that may or may be far from the speaker. As the number of objects that are not close to the speaker would surpass those that are close, there is nothing about this argument that can be used to evaluate the appropriate use of referring forms.

A second argument laid by Strauss against the traditional distance-based distinction is that it represents a static model, in which the speaker forms the center and demonstrative choice follows from the referent's 'relative distance from the speaker as the governing factor over demonstrative choice, with no other factors appearing to come into play' (p. 140). This claim is not so clear, as it is difficult to understand what 'relative distance' can mean if the only distance that matters is from the speaker to the referent. If the distance model does in fact only factor in how far a referent is from a speaker, then there would be no other distance by which this could be considered relative.

Strauss observes that 'this' is often used, e.g., (6), when referring to an object with an accompanying physical gesture, e.g., pointing, raising object up, moving object toward hearer.

(6) [History lecture, teacher: Koch]

Koch: Yeah. The border states—specifically [*pulls down map*] we've looked at this before, but which border states do you imagine he'd be particularly concerned about? Which two? What's this state? [*points to the area on the map*] (p. 141)

The form "this" is argued to be associated with new, non-shared information and is used to bring a referent into the consciousness of the hearer and, as evidenced by the use of physical gestures, is strongly associated with the "here and now". What is not clear about this explanation is what "here" and "now" mean in this model, if not the rewording of a speaker-centered framework. Another concern with the above explanation for 'this' is that counter-examples are available in which the speaker is attempting to bring a referent in the hearer's consciousness and high focus, yet "this" is a less felicitous form than "that," presumably because of the nearness of the referent to hearer, e.g., (7).

(7) [*A is in the audience waiting for B to give a speech, who is standing visibly off-stage. A calls B's cell phone just before B goes on stage*]

B: Hello, B speaking.

A: Why are you wearing a watch? It looks tacky.

B. Well, Gil's watch wasn't working.

A: I wasn't talking about Gil's watch. I'm talking about <u>that</u>/??<u>this</u> one you have on!

The perceived relative acceptability of "that" in (7) would be consistent with the distance model but not the gradient focus model, as it should be felicitous and in fact preferred to use "this" to refer to the watch in question, a high-focus object. Of course, such acceptability decisions would be treated well with judgments accumulated from native speakers, but invented examples are useful at least to frame the discussion for the time being.

The availability of counter-examples to the gradient focus model is worrisome. In (8), an example comes from my own son, Elias, then four years old, when he was being put to bed. I had just turned off his bedroom light, as well as the light in my own bedroom.

(8) Papa, can you turn back on <u>that</u> light?

There was no doubt in my mind that he was asking for me to turn back on the light in my room. If he had asked that I turn back on “this light,” by contrast, I would expect for the light in his room to be the intended referent. To this author, there is no easy means to explain the recognition of a referent between the use of “this light” and “that light” in the context of (8) with the concept of focus or redirection of attention. The question was not embedded in a larger conversation from which one can track the focus in discourse. It makes no sense in this scenario to compare the importance of the intended referents, nor can one easily differentiate between them according to new information versus shared information.

The above examples cannot pretend to discredit a framework based on solid corpus research and across language families, but they do give caution to a claim that attention or focus alone can differentiate the expectations for demonstrative forms.

1.4. The role of control in demonstrative form selection.

Langacker (2002, 2009) proposed that the concept of a *control cycle* can be usefully applied to several domains of language use. A control cycle describes a process of an agent starting at a baseline relaxation stage, upon which some new entity presents itself and causes tension in the decision for how to deal with the new entity, and afterward acts to and possibly succeeds to bring the entity under control, allowing the agent to reenter the relaxation phase.

The prototypical example is that of a cat and mouse. If a mouse were to wander into a cat’s vision, this would create tension, as the cat will very likely act on the sight of the mouse. It will pounce on the mouse, capturing, biting, and eating it, and thus release tension with its new control. At the risk of taking the example too literally, this author's initial objection to the presentation of the Control Cycle in this manner is that domesticated cats very often do not eat mice they catch but instead enjoy releasing a captive mouse only to recapture it. By relinquishing control and then attempting to regain it, an agent might therefore reevaluate its control in a changing world. According to this latter view of cats, the release of tension is more attractive than sustained control, in contrast to Langacker's (2009) claim that living creatures act with the goal of achieving and maintaining control. Langacker offers several examples to show control as a goal throughout life. On a social level, we experience tension when meeting new people and feel more control after getting to know them. Material possessions provide a feeling of control over our environment. As commonsensical as such examples are, it is overly simplistic to claim any one objective among all living creatures. Much human behavior, moreover, challenges an argument that control is a general goal, including such risk-taking behavior as recreational drugs and similar thrill seeking such as sky-diving.

Brovold and Grush (2012) applied the concept of the control cycle to demonstrative reference. They argue from the claims that (*i*) people seek, gain, and eventually lose control of entities, whether in a physical, perceptual, social, or attentional context; (*ii*) interlocutors in proximity are aware of each other's control relations; (*iii*) speakers are aware of the proclivities for various forms of control; and the knowledge of (*i*)-(*iii*) is the *control profile* of a speech situation, which can be defined as the "shared knowledge of all aspects of the control cycle, including: which entities your interlocutor is not paying attention to, is less likely to be able to perceive or grasp, what they used to, but no longer own."

Brovold and Grush develop their model for demonstrative use on corpora from sniper teams and on-line gaming forums, and they argue that the reason why speakers often assume that demonstrative use depends on spatial distance is because distance simply correlates with an individual's ability to control an object physically, perceptually, etc. The argument for language users employing demonstratives to reflect control relations, rather than spatial distinctions, is supported by examples where control, but not distance, can explain the expected use of a form. The authors offer an example from Italian, in which the so-called proximal demonstrative pronoun "questo" would be used in the utterance "Che cosa è questo?" (="What is this?") to refer to a small wound on the addressee's back, of which the latter might not be aware. Although the referent is close to the hearer, in fact, is part of the hearer, it is felicitous to use the proximal form in this situation where the speaker and not the mildly-wounded hearer can perceive the referent. The authors mention that spatial distance initially might seem to explain the acceptability of referring to a wound on someone's hand as "that" (or "quello" in Italian) when a few meters away from that person, but it is actually the speaker's restricted means of controlling the referent from a distance that drives the demonstrative form selection. Even from a distance of a few meters "this" could be felicitous if the speaker could (nearly) touch the referent with a stick. A recent example from playing with my sons was watching my son, Oliver, mistakenly dig in the pocket of the pants I was wearing for a hidden toy, at which point I said with the goal of being unhelpful, "I wouldn't look in that pocket." Although the pocket was part of the clothing I was wearing and thus by any measure proximal, I used the so-called distal form to specify the pocket where my son was searching, while I did nothing. Alternatively, one can explain the above example from Italian and the example with my son, Oliver, as cases where the referent was 'activated' by the speaker, thus undermining any expected role for relative distance (Gundel et al., 1993).

Perceptual control can also influence form selection. Brovold and Grush give the example of asking someone looking at something inside a box not in the speaker's view: "What's that?" Although the speaker may have been equidistant from the box as the hearer, the speaker could not see the referent of interest and thus would not ask "What's this?" If, however, the speaker sees something of interest inside the box and presumes the hearer cannot see it, it would be appropriate to say "Look at this" (or ask "What's this?"), but not "Look at that."

I propose a similarly expressed account of demonstratives, the form of which I propose can be predicted in English and Japanese from an ordered set of *control spaces*, spaces recognized according to some constraints (physical, linguistic, memory) and that allow for differing degrees of interaction with the referent. For what remains of this chapter, I will argue for how the concept of control spaces can explain demonstrative use in English and Japanese and, acknowledging the similarity to earlier control models, will try to highlight differences from Brovold and Grush's (2012) framework; show how control spaces better handle examples that are problematic for other models; and lay out predictions based on the current proposal.

The argument given here for a language user's negotiation of control spaces follows from a basic view of the function of language. A faithful representation of the world, according to this model, would be both excessive and uninteresting. It is argued here that the environment is processed according to the possibilities for interaction that are appreciated: to hold an object, to observe some behavior, to direct one's audience toward something, to wait for an object to approach the speaker, etc. A similar view of interaction with the environment can be seen, for instance, in such fields as robotics (Brooks, 1990, 1991), developmental psychology (Thelen & Smith, 1994), visual processing (Gibson, 1979; Nöe & Thompson, 2004), and gesture studies (Iverson & Goldin-Meadow, 2005; McNeill, 1992). According to an embodied view, language serves to manipulate that interaction, such as in the planning or interpretation.

The notion of control spaces is intended to avoid the binary categories of in-control versus not-in-control suggested by the control cycle (Langacker, 2002, 2009). Structures that may be physical, social, linguistic, or of other material are thought to constrain a "space," with which the interlocutors perceive or imagine some potential interaction. When referring to a physically-appreciated entity, the control space can be constrained by touching the referent, pointing at it, looking at it, or affecting it by other means. Although I will refer to control spaces as countable spaces that represent the action possibilities between an agent and an entity, when the relationships of many entities are considered simultaneously, we could envision the intersecting action possibilities as a *control field*, to account for the universal phenomenon of the control space between one agent and some entity affecting the control space between another agent and that entity and possibly even for the emergence of a group effect.

Control spaces are to be compared with each other on the estimated possibilities for interacting with an entity in relation both to interlocutors present and alternative referents. A minimal level of interaction could be considered one's distance to the entity. Other affordances would be signaled by the use of eye gaze, a pointing gesture, and touching or handling an object. Language and speech structures can also suggest degrees of possible interaction with the use of syntactic position, phonological or morphological stress, among other possibilities. The ranking of the control spaces constrained by the above behaviors will be treated empirically.

Of note, the action possibilities are imagined and may vary in their realism. For instance, if someone sees a get-away car speeding off, the action possibilities in reality might be limited, and the speaker might lament the impossibility of catching up with the car. An acknowledgement of the difficulty achieving that goal is only meaningful because the speaker can imagine an alternative scenario where she does catch up with car. For this reason, a car speeding away would likely be identified as “that car,” rather than “this car,” as “this car” suggests an imagined control space whereby the speaker could entertain the possibility of controlling the car more than is currently realistic.

The use of a demonstrative suggests more one control space. The selection of a form in some languages, such as English, depends on a ranking of the control spaces by their estimated action possibilities. The action possibilities may differ between a speaker and audience. If a woman picks up a telescope with her hands, we would rank her control space regarding it as a referent higher than that of a man watching her hold the telescope. The highest ranked control space will be the one for which the speaker appreciates the most action possibilities for herself (i.e. the first or 1’ control space) in relation to that entity. The 2’ control space will be the next highest ranked for action possibilities for the speaker and will often be the highest ranked control space for another agent, such as the hearer. The highest control space for the hearer is thus also a control space for the speaker, considering affordances indirectly available to the speaker through the hearer, but an inferior one to imagined action possibilities available directly to the speaker. The 3' control space, the next highest ranked, might be recognized between another agent not in the present discourse and the referent or simply between a noticeable landmark and the referent. In more complex demonstrative systems, the ordered array of control spaces could be expanded into a matrix, such that arrays could be ranked by other criteria to indicate the affordances to the interlocutors than the speaker's personal control. For example, in several languages spoken in mountainous regions (e.g., Lahu, Khasi, Byansi in the Himalaya, Lezgian in the Caucasus), the referent's vertical position (down, up) would suggest separate possibilities (Diessel, 1999).

The proposed framework describes form selection on the basis of action possibilities, rather than actual control. The example of a man handcuffed behind his back is a rare scenario in which the speaker could be physically touching the referent and yet, in addressing his liberator, use either the so-called distal form by saying: "Would you unlock <u>that</u>?" in recognition of his inability to free himself or "Would you unlock <u>this</u>?" in response to the illusion of control by touching an object. Imagined action, rather than actual control over an object, therefore drives the ranking of control spaces.

As the control spaces, in their simplest case, can be ordered according to the level of control possible or imagined by the speaker, the control spaces can be encoded by expected terms according to their rank order, e.g., in English, "this" for the first control space, "that" for the second. For Japanese, "ko" would be applied to the first control space, "so" to the second, and "a" to the third.

We can now consider how demonstrative forms in each language match up with various spatial configurations as an illustration of the use of control spaces to describe demonstrations, Fig. 1.2. Fig. 1.2 includes a man who is speaking to a woman about a flamingo. The rows are to be read as including only one flamingo at a time, to be chosen from one of the dotted cells.

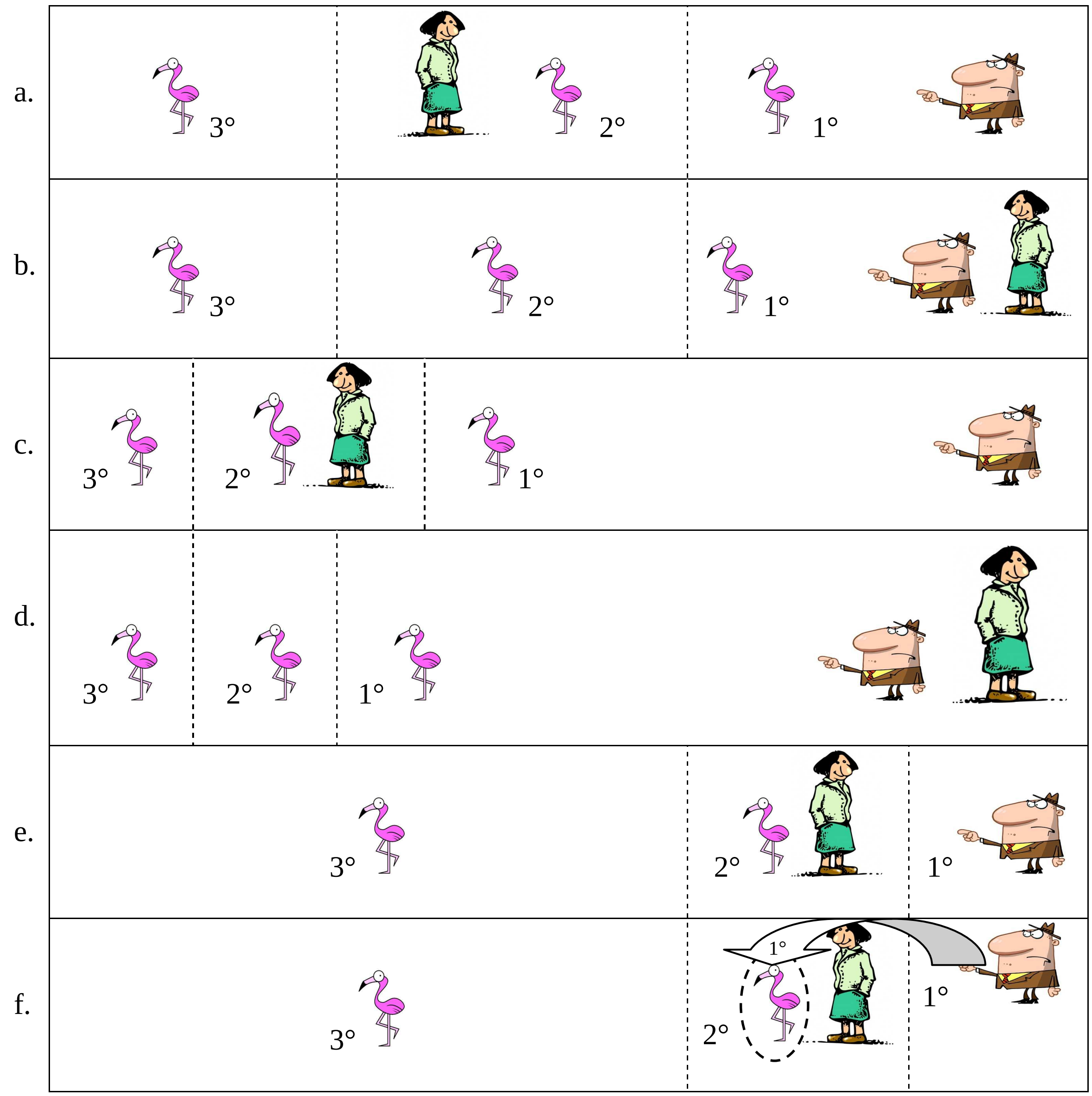

Fig 1.2. Schemata for control spaces secondary to discourse configurations

In Fig. 1.2a, the speaker and hearer have shared gaze, and neither is able to touch a referent from their position. The only way to decide the control space of the speaker is to evaluate his relative distance to the referent. The flamingo nearest the speaker lies within the highest-ranked control space. The flamingo nearest the hearer lies in his second highest-ranked control space. A flamingo beyond the speaker and hearer occupies a third highest-ranked control space. The first and second control spaces in Fig. 1.2a are meant to coincide with the findings in Coventry et al., (2008), as well as the intuition behind examples (1) and (2) above. We now can explain why the use of "this" in (2) sounds derisive, as it suggests the speaker is assuming control from the hearer regarding his own shirt.

If the speaker and hearer face the same direction, (Fig. 1.2b), then the control spaces are divided according to distances from the speech context to indicate the relative difference in action possibilities presumed for an object. Thus, an object might be considered within the first control space if the speaker intends to approach it, and it might be considered within a lower ranked control space if the speaker means only to note the object or is dismissive, reflected in example (4) above, and does not intend much further interaction with it. Fig. 1.2b is intended to reflect the intuitive difference between asking, from the curbside, "What do you think of <u>this</u> house?" versus "What do you think of <u>that</u> house?" discussed above. Notably, if an entity is referred to with a form associated with a lower-ranked control space, such as in Fig 1.2b, this implies the existence of an imagined higher-ranked control space. Referring to a house neutrally as "that house" without the suggestion of further action implies another, more primary control space, where one could act upon the referent, e.g., sign a purchase agreement for a house.

If the hearer is not attending in the same direction as the speaker, (Fig. 1.2c) and (Fig. 1.2d), then the speaker might extend the primary control space up to the hearer, as the interlocutors' relative distances to the referent will matter less if the hearer is not aware of the referent's location and thus cannot act upon it. A referent within the hearer's gaze would be in the secondary control space, and referents outside the hearer's proximity or gaze would be in a more remote, tertiary control space. Fig. 1.2c is intended as an analogy to Brovold and Grush's (2012) example of touching a wound on the hearer's back, yet within the speaker's control space due to the speaker's instrumentation of the wound with a stick. Fig. 1.2c also reflects the intuition in example (3), where a redirection of the hearer's attention negates the importance of physical distance. Fig. 1.2d illustrates also how one can point out a very distant object, e.g., "this galaxy!," and still treat it in the primary control space, because of the hearer's attention being elsewhere. Example (6) also involves a redirection of attention and is schematized in Fig. 1.2c and Fig. 1.2d.

The extension of the primary control space to a hearer attending elsewhere does not seem to work as well if the hearer blocks the path, with the exception of the speaker creating a path toward the referent. If the hearer lies between the speaker and referent, then the object will typically either fall in the secondary control space or lower, Fig. 1.2e. Fig. 1.2e also is meant to schematize the control spaces motivating the use of "that" to bring the hearer's attention to his own watch or a nearby lamp in (7) and (8). A situation can be imagined in which one uses "this" to refer an object behind the hearer. Imagine a shopkeeper at a bookstore pointing a finger and saying, "<u>This</u> is what you're looking for," referring to a book behind the shopper. According to this author's intuition, the use of the proximal form in this example suggests that the shopper is somehow transparent to the shopkeeper, who, due to familiarity with the store, can "see" the book even with a person's body in the way. As Fig. 1.2f depicts, physical distance is not the only criterion to judge control space. If the speaker sees a path for himself toward a referent object behind the hearer, then the object could be assessed with higher action possibilities for the speaker than the hearer. An example from my own house is when my son asked me for “that book over there” from the comfort of his bed. When I did not know which one he had in mind, he said, with a huff, “this book over here,” as he got up to fetch it himself.

Fig. 1.2f is intended to show how even an object can be included in the speaker’s primary control space, despite the hearer’s relative nearness to the object, to reflect other means of controlling an object that are not physically apparent. In the case of the father owning a car that is nearer the son in (5), there is a bond between the owner and father that the interlocutors can appreciate that is not material. Societal connections are very meaningfully represented through our language choices, even though they are not available directly to our senses.

The demonstrative forms in English and Japanese are posited here to be associated with the control spaces schematized in Fig. 1.2 and listed in Table 1.1.

	1°	2°	3°
Japanese	こ “ko	そ “so”	あ “a”
English	this	that	That

Table 1.1. Demonstrative forms and associated control spaces

The various configurations of control space in Fig. 1.2 suggest many hypotheses, but this dissertation will be limited to three hypotheses, in order to separate the control space description of demonstratives from that of a distance model and attentional models.

As stated above, Hypothesis 1 is that users of separate languages will depend on extra-linguistic information, such as a pointing gesture, differently. Table 1.1 indicates that English does not differentiate the secondary and tertiary control spaces lexically, whereas Japanese uses the morphemes “so” and “a” for contexts comparable to when English speakers use “that.” Considering that Japanese employs more lexical means with which to constrain a control space, it is possible that English speakers will rely more on pointing gestures to restrict search space. This prediction would relate to the Mutually Adaptive Modalities Hypothesis (De Ruiter, 2006; Melinger & Levelt, 2005), discussed in Chapter 2, which predicts that gestures will supplement lexical communication in language-specific ways. Hypothesis 2 expects for spatial distance to affect demonstrative form, because spatial distance will correlate highly with action possibilities. Per Hypothesis 3, if the hearer is attending elsewhere than the area of the referent, then spatial distance should matter less, given the speaker’s greater affordances with the referent without the hearer’s attention. Table 1.2 separates the Control Space description from the distance and attentional models according to the three hypotheses, as they were presented in Sections 1.2 and 1.3 (thus, the Gradient Focus model will serve as the example of an attentional model).

	Hypothesis 1	Hypothesis 2	Hypothesis 3
Distance model		✓	
Attentional models			✓
Control space	✓	✓	✓

Table 1.2. The hypotheses and their fit among demonstrative descriptions.

We can see therefore that demonstratives have been treated in quite different ways by various theorists: as markers of distance, cues for attention and focus, correlates of the estimated cognitive state of the hearer, as well as the level of control between the speaker and object. In what follows in Chapter 2, I will discuss evidence from primatology and clinical and gesture studies to support the argument for control spaces as a viable description for how humans demonstrate in space. In Chapters 3 and 4, I will introduce behavioral and electrophysiological data to support the predictions from the Control Space framework. The advantage of using behavioral data is that we will have indices, such as reaction time, for the ease of cognitive processing. The event-related potential paradigm is used because it can offer detailed information regarding the temporal progression of violation detection, broad localization techniques on the basis of scalp topography, comparisons with other studies of semantic and pragmatic anomalies, and can be used to corroborate data obtained by other techniques. The ERP technique is thus important for its measurement of brain electrical activity to provide data that are not available by behavioral data alone and can serve as a preliminary study of how the brain processes referential language use in space. In Chapter 3, I will discuss an ERP study that offers evidence both for spatial distance as a factor in demonstrative selection (Hypothesis 2) and that spatial distance will be a less important factor if the speaker and hearer do not have a shared gaze (Hypothesis 3) (Stevens & Zhang, 2013). In Chapter 4, I will discuss an ERP study (Stevens & Zhang, 2014) that provides support for the view that reliance on a pointing gesture depends on the specific language (Hypothesis 1). In Chapter 5, I will review the findings discussed thus far in light of further possible investigation, such as in the contexts of autism and schizophrenia, and will provide a preliminary description of demonstratives as single-use names (or *disposable* names).

Chapter 2: Overview of gesture

From the forest to the crib to the din of the factory

2.1. Introduction.

The topic of demonstratives attracts theorists and researchers from a broad range of disciplines partly because demonstratives seem so primitive. This impression is reinforced by the co-occurrence of gesture with demonstratives, as gesture itself seems like a predecessor toward full-fledged symbolic language. In Chapter 1, we introduced the concept of *control spaces* as physical or metaphorical dimensions by which available actions toward a referent are assessed by the speaker. By this view, when using a demonstrative, we encode a basic categorization of possible interaction.

To evaluate the concept of control spaces, it will be helpful to review how gesture has been analyzed across species and developmentally. As discussed below, gesture and language appear to develop to influence one's environment. Both in non-human apes and among children, control of an object will drive communicative tools in various forms. The discussion to follow takes the view that demonstratives are closely associated with gesture use, but alternative viewpoints exist, such that demonstratives are considered primarily as terms correlated with mental states and secondarily terms employed in physical space (Gundel et al, 1993).

2.2. Gestures among non-human primates.

Gestures often resemble the expected actions that motivate their use (Arbib, 2002; Armstrong & Wilcox, 2007; King, 2004; Perlman, Tanner, & King, 2012; Tanner &

Byrne, 1996). Call and Tomasello (2007) have studied the use of gestures among great apes in indicating such potential behaviors as grooming, play, dominance, nursing, and sex that require social negotiation and found similarities between the sign and the potential action. Whereas non-human primates seem to have a fixed number of vocalizations, their gestures seem to be individually learned and flexible with their usage (Pika, Liebal, & Tomasello, 2003; Tomasello, George, Kruger, Farrar, & Evans, 1985). In studying captive apes, Savage-Rumbaugh and colleagues (1986) observed pygmy chimpanzees making twisting motions to request another to open a lid or a hitting motion to encourage another to crack nuts. The repertoire of non-human ape communication is far more restricted than humans, not only due to the lack of verbal language in the former but also in the expression of imaginative scenarios. Tomasello (2008) argued against non-human primates' capacity for iconic gesture, arguing that this presupposes a sophisticated theory of mind. He offers the example of the human gesture for sprinkling (invisible) cheese, a task unlikely to be observed among other apes, as a counterpoint to any view extending imitative human gesture use to other apes. Rather than delineate the capacity of one species from others so categorically, another approach would argue for a spectrum of *simulative complexity*, such that humans could entertain more abstraction than other species that may ground gesture more closely to the intended action (Perlman et al., 2012).

The pointing gesture is one that occurs spontaneously among humans but does not occur consistently among other apes in their natural environment (Racine, 2012). Although captive apes do point (Leavens, Hopkins, & Bard, 1996; Leavens, Hopkins, & Thomas, 2004), there is controversy what this behavior really means and how it relates to what humans do (Tomasello, 2006, 2008) and if chimpanzees could possibly have the social-cognitive abilities required to point at an object for the sake of an audience (Povinelli & Vonk, 2003). Chimpanzees have been observed to point to food out of reach and, unlike the typical use of the index finger among humans, have been seen to point with the whole hand (Call & Tomasello, 1994; Wilkins, 2003) and feet (Woodruff & Premack, 1979) and use lip pointing as is done in some human societies (Enfield, 2001, 2003). Because chimpanzee pointing often targets food, it could be tempting to call this "reaching," rather than pointing, but Leavens et al. (Leavens et al., 1996, 2004; Leavens, Russell, & Hopkins, 2005) have noted that this behavior does not occur when the chimpanzee is alone. Similar behavior has been observed among other captive apes, including bonobos, gorillas, and orangutans (Leavens & Hopkins, 1999), as well as monkeys in captivity (Blaschke & Ettlinger, 1987; Hess, Novak, & Povinelli, 1993; Kumashiro, Ishibashi, Itakura, & Iriki, 2002) and even dolphins (Xitco, Gory, & Kuczaj, 2001). It would be easy to discount behaviors among captive animals as uninformative by challenging the ecological validity of this observation, but this line of reasoning could be leveled equally against any descriptions of urban human beings (Leavens et al., 2005). Wild bonobos were once observed to point at human observers attempting to hide nearby (Véa & Sabater-Pi, 1998), wild non-human apes do not point with the same frequency as captive ones.

However, it should be noted that non-human primates do use gesture abundantly in the wild, just not to pick out a referent. For instance, male chimpanzees might indicate their sexual intentions towards females with body movements, rather than vocalization, so as to avoid signaling their interest to competing males (Hobaiter & Byrne, 2012).

Tomasello and colleagues have observed the use of raised arms to indicate an interest in play and poking behavior to get attention (Tomasello et al., 1997; Tomasello, Call, Nagell, Olguin, & Carpenter, 1994). The close association between the social intention and the communicative gesture among gorillas was catalogued by Tanner and Byrne (1996), such that gorillas used gestures that, to a human observer's eyes, demonstrate the actions they wanted other individuals to perform. Chimpanzees likewise can be argued to use two types of intentional gestures: "incipient actions" transformed into gesture through ritualistic use and "attractors" (Pika, Liebal, Call, & Tomasello, 2005). Incipient actions include, for example, a young chimpanzee that touches her mother's rear-end to signal that her mother lower her back to climb on. An example of an attractor would be males' use of *leaf-clipping*, whereby they make noises to attract the attention of females for sexual behavior.

There is no straightforward explanation for why apes raised in captivity point but ones in the wild typically do not, but a possibility is that the physical restraints of captivity lead to the spontaneous use of pointing in both chimpanzees and humans because of recognition of a *referential problem space*, "a set of related circumstances in which barriers exist to direct attainment of goals and therefore indirect means to attain those goals must be discovered or recalled then effectively applied" (Leavens et al., 2005, p.188). Whether one is chimpanzee or human, the use of pointing, therefore, can be thought of as an epigenetic strategy to interact socially with one's surroundings, initially with the basic goal of obtaining an object with the aid of another agent. Rather than depending on one's own body directly for a solution, a physically-restricted agent can exploit a *referential triangle*, including the sign, the referent, and the intended audience, to achieve a goal that would be either impossible or laborious through one's own effort (Butterworth, 2003; Leavens et al., 2005).

2.3. Gesture among human language learners.

The concept of non-human primates performing a referential gesture has been challenged on the basis of the complex social-cognitive abilities involved, although one-year old humans also point and understand pointing, without the supposedly necessary social-cognitive abilities, seen at least two years later (Behne, Carpenter, & Tomasello, 2005; Camaioni, Perucchini, Bellagamba, & Colonnesi, 2004; Liszkowski, 2006; Wellman, Cross, & Watson, 2001). Liszkowski (2006) mentions that human pointing may differ categorically from pointing in other species ontogenetically, such that it has been argued that human pointing is associated with symbol use (Werner & Kaplan, 1963) and language acquisition (Goldin-Meadow & Butcher, 2003; Wilkins, 2003). Pointing among humans shows a similar manipulation of shared attention, although the form of pointing varies across cultures, including the use of a finger, a whole hand, or a lip pointing (Wilkins, 2003). No matter the phenotype, Butterworth (2003) considered gestures always involve a learned association combining the signaler, an audience, and an object.

What has been referred to as deictic or referential gesture may form an important transition between non-symbolic and symbolic representation. Tomasello (2003) categorizes infant gestures into ritualistic, deictic, and symbolic: the first of which is not symbolic, the third is symbolic, and the second forms an intermediate form. Ritualistic gestures are ones that infants perform to stimulate a response, such as lifting their arms to get picked up, and Tomasello links such gestures with primate gesture use. Deictic gesture use differs importantly from ritualistic use, according to Tomasello, in its formation of a triad of signaler, interactant, and some other entity of some presumed attentional value, whereas ritualistic gestures involve only the signaler and interact. Tomasello's formulation applies a division between procedural gesture and attention-directing ones, but one may understand ritualistic gestures such as "arms up" as a means to redirect the interactant's attention to the signaler's own body as a third object.

As with non-human primates, there exists a degree of speculation in describing an infant's intention when pointing. Bates, Camaioni, and Volterra (1975) described an infant communication as following a development along three stages, following Austin's (1962) speech act theory: a perlocutionary stage, an illocutionary stage, and a locutionary stage. The communication at the perlocutionary stage would involve leading to some result as a by-product of the communication. At the illocutionary stage, communication is used with the intention to lead to some result. At the locutionary stage, communication serves to makes statements without an explicit intended result. Bates et al. argue that infant pointing is at the illocutionary stage, such that it expresses an intention, but in later work it was argued that babies use both *proto-imperative* pointing and *proto-declarative* pointing (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979). As the names suggest, proto-imperative pointing is used to indicate something that the baby wants done, and proto-declarative pointing serves to focus the audience's attention on some interesting state of affairs.

Imperative pointing and declarative pointing have been treated as distinct behaviors that require different cognitive prerequisites. Imperative pointing, that is, indicating one's want with gesture, has been treated as a sort of ritualized failure to reach an object (Vygotsy, 1978), but pointing and reaching for an object have been shown to be unrelated developmentally (Masataka, 2003). Reaching and imperative pointing differ importantly in that, among mature users, pointing is performed only with another individual present and is therefore a social act, such that it manipulates another person's body. Reaching, on the other hand, is a physical act that may be executed without consideration of others. When pointing declaratively, the agent is picking out an object to manipulate another person's attention. This sort of pointing appears more sophisticated, in presupposing the user understanding the audience's attention state.

To support the view that declarative pointing requires more advanced cognitive abilities, it has been reported that non-human apes and autistic children do not point declaratively, although they do so imperatively (Liszkowski, 2006). The lack of declarative pointing in the second year of life serves as an essential diagnostic criterion for autism (Baron-Cohen, 1996). Declarative pointing, on a related note, seems to emerge in line with the emergence of productive language (Carpenter, Nagell, & Tomasello, 1998).

The age that infants start pointing imperatively differs according to reports. Desrochers, Morissette, and Ricard (1995) mark such behavior as early as 15 month olds, who are sensitive to their audience's gaze. Liszkowski (2006), however, provides evidence for 12 month olds pointing imperatively. Furthermore, infants have been found to point even when they are alone in a room, challenging the notion that pointing behavior depends on gaze recognition (Delgado, Gómez, & Sarriá, 1999). Moore and D'Entremont (2001) found that 12 month olds point repeatedly at an object without accounting for their audience's gaze. Povinelli and O'Neil (2000) pose complications with the distinction between imperative and declarative pointing, remarking that these pointing functions have been defined differently according to researcher in terms of gesture identification and whether alternating gaze accompanied the pointing. These authors raise the question of whether tracking of gaze follows a separate physiological system from pointing that cannot work as distinguishing feature of among pointing functions.

In light of reports that challenge a neat developmental model of gesture use, we should exercise caution in our interpretation of imperative and declarative pointing. Firstly, any talk of development of kinds of pointing suggests that the mature forms exist, but what we call ‘imperative’ and ‘declarative’ may be more conveniences based on prototypes than natural kinds. Examples from language are abundant, e.g., the declaration “There’s no salt” may be understood as an indirect imperative to provide salt. Even a more abstract declaration, such as “New Mexico has the highest number of Spanish speakers per capita for a U.S. state,” carries the expectation that the hearer begin to assume the speaker’s knowledge. In other words, any declaration *S* can be understood equivalently as an imperative “Know that *S*.” Interrogatives can be treated analogously. Referential pointing can similarly be thought to have an imperative quality, e.g., pointing at a couch in a furniture store carries the expectation that the hearer consider this, perhaps as a purchase.

If all behavior can be thought of as manipulation of one’s environment, all utterances and gestures can be treated as an imperative, commanding the audience to one’s attention. The development of referential pointing, eventually leading to words, shows a gradual release of control, as children begin making reference by showing an object, transition to handing it over, and finally pointing (Volterra & Erting, 1990). As a human matures, language and gesture appear to operate from a common planning function (McNeill, 1992). Evidence for the common production system for gesture and language is found in mirror neurons, cells responsive to the observation of others performing an action, being activated in frontal-parietal areas when perceiving or performing both hand and mouth movements (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti & Arbib, 1998).

The ontogenetic transition toward less physically restricted forms of control in object reference is mirrored in language development. The use of deictic language is more prevalent than representational descriptions among young children, but this pattern reverses as language users mature (Pizzuto & Capobianco, 2005). Several studies have shown that referential gestures are the most prevalent among children 16-20 months of age, where the most commonly used gestures were used with words to name an object (e.g., saying “flower” while pointing at one) or comment on it (Caprici, Iverson, Pizzuto, & Volterra, 1996; Pizzuto, 2002; Volterra, Caselli, & Capirci, 2005). Even after children develop a more sophisticated lexicon and can express themselves with words only, they continue to point to pick out an object with their bodies (Rodrigo, Gonzalez, de Vega, Muneton-Ayala, & Rodriguez, 2004).

The progression of gesture and language use among young humans cannot be described in sure terms, as individual variation can easily upset a developmental model. However, gesture and language acquisition both can be viewed as the obtainment of skills to control one’s environment at a distance. Pointing among infants seems to emerge as a means to use one’s audience as a tool to satisfy an immediate desire and eventually includes declarative or referential pointing as a strategy to manipulate the audience. As noted above, the act of reference itself reflects how the learner develops less direct ways to control their surroundings: first by holding up the object of interest, then handing it over, and finally pointing at it (Volterra & Erting, 1990). The manipulation of one's environment by recruitment of one's audience seems to drive not only gesture development but also complements the view that a demonstration of an object includes an assessment of the action possibilities available with which to control it.

2.4. Language and gesture among mature users.

The relationship between gesture and language use is not yet clear, as it is possible that gesture serves to supplement concepts not expressed in language or that it accentuates concepts constrained by the possibilities set by language, among other possibilities.

One framework conceptualizes language and gesture as originating from a common productive source, and the structural constraints of a particular language will therefore affect how physical gesture occurs. By this view, syntactic and semantic structures of languages represent events in such a way that affects the use of co-speech gesture (Halliday & Hasan, 1979; Kita & Özyürek, 2003; Kita, 2000; Lyons, 1977; McNeill & Duncan, 2000; Özyürek, Kita, Allen, Furman, & Brown, 2005; Quirk, 1979). Theoretical arguments have emerged from this evidence that conclude that utterance generation includes an interaction of language and gesture, stated as the Interface Hypothesis (Kita & Özyürek, 2003) or Growth Point Theory (McNeill & Duncan, 2000).

McNeill and Duncan (2000) for instance, studied the use of gesture and speech in English and Spanish, languages that encode verbal actions differently for manner and path. Manner is generally less explicitly expressed in Spanish than in English. The study presented English and Spanish speakers with excerpts from a Sylvester and Tweety cartoon that included motion events and found that Spanish speakers are more likely to express the manner of an event with a gesture, more so than in their speech, whereas English speakers expressed both manner and path in their speech and their gestures often downplayed the manner of action. The authors treat specific languages and gesture systems as modes of "thinking-for-speaking," to which end gestures are considered "material carriers" that will take on language-specific forms. Gestures, in this view, are used more frequently when the context differs appreciably from the thought to be expressed, because the gesture will further contribute to the meaning. Gestures, in fact, have been observed to be used more elaborately when there are discontinuities or less predictable uses of speech (McNeill, 1992).

The Interface Hypothesis claims that the linguistic representation and the gestural/imagistic representation of an event interact during speech (Kita & Özyürek, 2003). According to this hypothesis, when formulating one's speech, gestures encode both the imagistic aspects of a representation and the conceptualization of an event. If some component of an event is not expressed in speech, it will not be expressed in gesture. If speakers of languages employ different syntactic frames to express a concept, then the gestures will accordingly. If a speaker's language does not lexically encode a concept, then there will not be an analogous gesture. For example, English has a verb "swing" that has no equivalent in Japanese and Turkish, neither of which have a concise means to express movement along an arc. Kita and Özyürek (2003) found that Turkish and Japanese speakers were more likely to gesture a swinging concept with a straight movement of the arm, whereas English speakers swung the arm along an arc trajectory. As a second experiment, the authors compared English, Turkish, and Japanese speakers' expression of the path and manner of event. Japanese and Turkish do not have the capacity as English speakers to express manner and path in the same clause, e.g., "He rolled down the hill." Japanese and Turkish speakers would instead express this concept by separating the manner and path, e.g., "He descended as he rolled." In support of the Interface Hypothesis, it was found that Japanese and Turkish speakers were more like to make two gestures when expressing this concept, and English speakers used only one.

An alternative view to the Interface Hypothesis is the Mutually Adaptive Hypothesis (De Ruiter, 2006; Melinger & Levelt, 2005), which expects for gestures to complement speech in settings where speech is less effective, e.g., when there is loud ambient noise. Gesture is used less when speech is more effective, such as over the telephone. Furthermore, speakers gesture less when their descriptions of an object are more detailed (Melinger & Levelt, 2005). The Mutually Adaptive Hypothesis was developed in the context of a single language, unlike the Interface Hypothesis, but one can extrapolate the Mutually Adaptive Hypothesis to predict a trend by which languages with more lexical means to constrain a search space require fewer extra-linguistic means.

The above hypotheses differ in how they explain the association of gesture with language, and it is admittedly quite difficult to understand how language and gesture are produced, but the frameworks of language and gesture production above all do describe language-specific gesture use. We expect for the parameters of a given language to influence gesture in a demonstration if we consider these communication systems as combinatory tools by which a space of action possibilities can be designated. It would be useful to consider what the above hypotheses might suggest regarding the sensitivity of language users to a pointing gesture in correlation with the richness of that language's demonstrative system. In the following two chapters, we will investigate how electrophysiological components in an EEG signal reflect language users' intuitions regarding demonstrative use in space and with such expected parameters of a control space as gaze and gesture.

The subject population will include both native English and Japanese speakers. The questions of interest are whether the presentation of demonstratives in varying visual contexts will elicit a previously reported brain response or a new neurophysiological finding. One probable candidate response would be an N400, as this ERP had been observed following semantic or pragmatic challenges. It is not clear from previous N400 studies if similar findings would occur with spatial demonstrative use. Another also is unclear from the literature is the role of the P600 ERP, the positive deflection at approximately 600 ms, regarding semantic processing. As there have not been ERP studies reported about demonstratives, it is moreover unknown how speakers of languages with different demonstrative systems would respond differently. Some components of a natural demonstration will serve as variables in the ERP studies, including the presence of a pointing gesture and shared gaze.

Chapter 3:

The role of shared gaze: An ERP study of English demonstratives

3.1. Introduction.

The concept of control spaces as a means to explain demonstrative use should be evaluated by its predictions. A control space is formed by constraints, physical or symbolic, and allows for agents to negotiate how they might pursue action with an object. One way that a speaker could distinguish possibilities for action would be to categorize whether an object is more easily controlled by the speaker or her listener. All else being equal—both agents are able-bodied and are facing the object, etc.—one way to evaluate each agent's potential for higher control of an object could be simply their relative distance to the object. If an object is close to the speaker, we might categorize this space differently than if the object is close to the hearer, as the hearer would then be able to exert more control over the object. However, even if the hearer is close to the object, we would not necessarily expect for her to exert the most control over an object if she is not attending in the direction of the referent. The notion of control spaces thus leads to two initial predictions, corresponding to Hypotheses 2 and 3 in Chapter 1. Firstly, language users are sensitive to relative distance in the acceptability of a demonstrative form. Secondly, the sensitivity based on relative distance will be reduced if the hearer is not attending in the direction of the referent.

In order to consider these predictions, an experiment was performed that tests language users' responses to visual scenes accompanied by different demonstrative expressions. The experiment tested both behavioral and electrophysiological responses to simultaneous presentations of a picture and an auditory stimulus, to which participants either expressed that they agreed or disagreed that the combination would be an expecting pairing. The experiment described below was published in the Stevens and Zhang (2013) paper, “Relative distance and gaze in the use of entity-referring spatial demonstratives: an event-related potential study.”

As discussed in Chapter 1, there are various accounts for demonstratives in English and other languages. Traditional accounts of English demonstratives have emphasized distance from the speaker as a distinguishing feature of demonstrative forms (Halliday & Hasan, 1979; Lyons, 1977; Quirk, 1979). Other accounts emphasize the role of attention (e.g., Diessel, 2006; Fillmore, 1971; Himmelmann, 1992). Accounts that use similar concepts of attentional state, cognitive status, and focus also have been proposed (Fillmore, 1971; Gundel et al., 1993; Hanks, 1992; Strauss, 2002). The model proposed presently, that of control spaces, seeks to build on the strong points of both a distance model and an attentional model. It is argued here that relative distance to the speaker or to another participant is relevant, as distance will factor into the action possibilities available to the respective agents. However, relative distance as a distinguishing feature can be easily superseded by attentional factors, such as redirection of the hearer’s gaze. According to the control spaces model, a referent that is near the hearer can still be labeled with the so-called proximal demonstrative “this” if the hearer is not looking in that direction, as the speaker would be considered to exert more control over the referent in that context despite the greater physical distance because of the hearer’s attention elsewhere.

Despite the linguistic investigations into demonstratives, we still lack a basic understanding of the mental processes underlying demonstrative use in spatial reference, including the neural mechanism to our intuitions for their usage. Psycholinguistic evidence gives some support to the view that demonstratives serve to demarcate distances from the speaker (Coventry et al., 2008). It has been noted earlier, however, that distance cannot alone explain demonstrative use, as there are many other extra-linguistic factors to consider, including gaze and the use of a pointing gesture. All cultures, for instance, use gesture, and all languages include demonstratives used with pointing, and all languages have demonstratives associated with pointing (Diessel, 2006). Moreover, the finding of distance factoring into demonstrative form choice did not consider the role of distance to a hearer in the Coventry et al. study.

In order to address the above questions in a novel matter, we used the event-related potential (ERP) technique in combination with behavioral measures such as acceptability judgment and reaction time. The ERP technique makes use of averaged EEG recordings time-locked to stimuli and provides a measure of temporal description of neural activity. Different experimental paradigms can evoke recognizable patterns of EEG signal. As an example, a signal peak of negative potential at approximately 200 ms after a stimulus, called the *N200*, detects a mismatch of information (Sutton, Braren, Zubin, & John, 1965), even for language tasks (Schmitt, Münte, & Kutas, 2000). A possibly related ERP component, the *N400*, has been found to register semantic information or contextual incongruities (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980, 1984; Osterhout, Willems, Kita, & Hagoort, 2002). If, for instance, a subject was presented with the sentence “I always put on my socks before my car,” we would expect to observe a larger negative peak at 400ms after the onset of the final word than would be recorded after the presentation of the sentence "I always put on my socks before my pants."

Although no ERP studies previous to Stevens and Zhang (2013) have been reported that compare the use of demonstratives to pick out objects in space, several ERP studies have taken up the question of the interpretation of gesture use (Cornejo et al., 2009; Holle & Gunter, 2007; Kelly, Kravitz, & Hopkins, 2004; Kelly, Ward, Creigh, & Bartoletti, 2007; Wu & Coulson, 2005, 2007). Kelly et al. (2004) elicted an N400 effect by pairing hand gestures indicating spatial dimensions with words that connote spatial concepts (e.g., tall, short, wide), if the gesture and word did not match (e.g., pairing the word "short" while gesturing *wide* by spreading out one's hands). An N400 effect has also been produced when a gesture did not correspond with the interpretation of an ambiguous word (Holle & Gunter, 2007). In that experiment, participants were given ambiguous information, such as "Everybody was impressed by Sandra. She controlled the ball," where "ball" could mean either a setting for dancing or an object used in a game. An accompanying gesture suggested one of these two meanings, but if the verbal material that followed indicated the other meaning of the ambiguous word, a late negative waveform was observed. Cornejo et al. (2009) paired gestures with metaphorical speech. For example, the participant might hear (in Spanish) "Those young people are giraffes." In the incongruent condition, the gesture accompanying the verbal stimulus "giraffes" is a hand held at the waist indicating short height, and this pairing yielded an N400 effect.

Özyürek, Willems, Kita, and Hagoort (2007) compared the time course of ERP responses to incongruent gestures with those to incongruent speech and found that semantic interpretation of gestures corresponded with the time course of the interpretation of speech, suggesting a similar neural processing mechanism. These studies of gesture use support the view that gestures contribute to the interpretative process, in that bimodal information (i.e. speech and gesture) is integrated into a single conceptual representation (McNeill, 1992). An opposing view contends that gestures are epiphenomal effects that do not contribute significantly to the interpretation of information (Krauss, Dushay, Chen, & Rauscher, 1995). It is not straightforward how this debate informs the interaction of gesture with demonstrative use. Demonstratives rely on extra-linguistic cues such as gesture more so than words with more conceptual meaning that may uniquely pick out an object in a context without additional, non-linguistic behavior (e.g. the short glass vs. that glass).

From the above studies of gesture comprehension, we might consider an N400 component to be the most likely effect observed from incongruent spatial demonstrative use, given the similarities of gesture and demonstrative use, as well as the repeated finding of an N400 effect from other contextual surprises, such as the strangeness of the image of a man in a business suit standing on one leg in the desert (Proverbio & Riva, 2009) or simply saying “Dutch trains are white” to subjects in the Netherlands, who would all know that Dutch trains are yellow (Hagoort, Hald, Bastiaansen, & Petersson, 2004).

The *P600*, an ERP that shows a positive wave deflection at approximately 600 ms post-stimulus was previously thought to reflect only syntactic anomaly (Hagoort, Brown, & Groothusen, 1993) but has been observed in response to special cases of semantic incongruity (Kim & Osterhout, 2005; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Nieuwland & Van Berkum, 2005). Kuperberg (2007) argues that a P600 effect will be observed more likely than an N400 when there is a violation of semantic verb-argument relationships. For example, Paczynki, Kreher, Ditman, and Holcomb (2006) were able to differentially yield an N400 effect and a P600 effect based on the animacy requirement of a verb. When a subject heard the phrase "At long last the man's pain was understood by the ...," an unremarkable completion to this sentence would be "doctor." A doctor is not only capable of understanding pain, being animate, but also is contextually appropriate, as people typically discuss their pain with health care professionals. If, instead of "doctor," the final word had been "violinist," an N400 effect was observed, because violinists are not usually expected to hear of someone's pain. When the final word was instead "pens," then a P600 effect was instead seen. Pens do not satisfy the animacy requirement to understand something, which makes the sentence difficult to process in a way that differs categorically from hearing of a violinist understanding one's pain, an unusual but not impossible situation. A P600 effect is similarly observed in other experiments where the NP does not satisfy a requirement of the verb, e.g., "The hearty meal was devouring...," as devour requires an animate subject, in comparison to the acceptability of it having an inanimate object: "The hearty meal was devoured..." (Kim & Osterhout, 2005). An unexpected use of a demonstrative, such as the above example of using "that" to refer to the shirt one is wearing, might be considered more likely to evoke an N400-like phenomenon, since the unexpected use of a demonstrative form within a spatial context still has meaning, unlike pain being understood by a pen. However, given the dependence of demonstratives on additional cues to successfully pick out an object, it is also possible for a P600 effect to be observed, if the structure of a demonstrative form and extra-linguistic cues are not compatible.

In addition to the above work on gesture using EEG, several other studies have used fMRI (functional Magnetic Resonance Imaging) to localize cortical regions responsible for processing gestures (Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009; Green et al., 2009; Xu, Gannon, Emmorey, Smith, & Braun, 2009). For instance, Green et al. (2009) found that when participants observed gestures that matched accompanying speech, occipital regions were activated, whereas parietal and posterior temporal brain regions were activated if the gesture was unrelated to concurrent speech. Gesture comprehension and spoken language were also found to map to the superior and inferior temporal cortices, respectively (Xu et al., 2009). The left anterior frontal gyrus and bilateral posterior superior temporal sulcus responded more to speech when a gesture accompanied the auditory information. The right inferior frontal gyrus responded to the semantic processing of hand gestures (Dick et al., 2009). The left inferior frontal gyrus activation for semantic processing was reported in several other studies (Grindrod, Bilenko, Myers, & Blumstein, 2008; Moss, Rodd, Stamatakis, Bright, & Tyler, 2005; Zhu et al., 2009).

The ERP study discussed in this chapter is the first project to investigate the neural mechanisms behind demonstrative use for objects in space (Stevens & Zhang, 2013). The study measures the semantic congruency/incongruency effect of a visual scene and an auditory phrase that contained an English demonstrative (“this one” or “that one”). There were two primary goals to the project. The first goal was to investigate whether an N400 effect, seen in previous ERP work on semantic processing, would be elicited from the presentation of demonstratives in unexpected ways. Moreover, there would be a question if differences would emerge regarding the time course, the waveform morphology, and the scalp distribution in a study of demonstratives with a visual context, as compared to studies with verbal stimuli only. A second goal sought to localize those brain regions that respond with the judgment of acceptable demonstrative use in space. The hypothesis related to the first goal is that an N400 effect would be observed, similarly to gesture ERP projects, when subjects are presented with unexpected pairings of demonstratives and visual scenes. Given the newness of this experimental design, it is difficult to specify the hypothesis further, but we might expect for the latency of the ERP component to increase in light of the complex task of integrating relative distance and joint gaze in conjunction to a verbal stimulus. Again, we would expect for the acceptability judgment of spatial demonstrative use to be similar to that in the N400 studies and would therefore expect a dominant posterior negativity scalp distribution (e.g., Johnson & Hamm, 2000) and specifically a parietal cortical response in line with semantic processing studies that have made use of fMRI (Chou, Chen, Wu, & Booth, 2009; Grossman et al., 2003).

3.2. Methods.

3.2.1. Participants.

Participants include sixteen native English speakers, 1:1 ratio of male:female, recruited by advertisement. The average handedness index of the subjects was +0.9 (right-handed), according to the Edinburgh handedness inventory (Oldfield, 1971). The average age of subjects was 25, ranging between 21-34 years of age. All subjects had normal hearing, normal or corrected vision and no known history of color-blindedness, speech, language, or hearing disorder. The protocol approved by the Human Subjects' Protection Program in coordination with the Institutional Review Board at the University of Minnesota was used for the purpose of obtaining consent from subjects. The subjects were each paid $30 for their time. The data from two subjects was removed from final analysis, because insufficient trials were available due to excessive blinking and other muscular activity.

3.2.2. Stimulus material.

To produce the auditory stimuli, an AT&T text-to-speech was used (http://www2.research.att.com/wttsweb/tts/demo.php). The stimuli for the demonstrative phrases included “this one” and “that one” in a female voice. The use of only a female voice was chosen to avoid confusion regarding the role of the characters. The samples were digitally edited, and they were resampled at 44.1 KHz with the use of Sound Forge 9 (SONY Corp.) software, and the root mean square intensity levels were normalized and equalized.

The visual stimuli were created with the use of Second Life software, a program with which virtual world scenarios can be generated in 3D (http://www.secondlife.com). The scenes all included a woman pointing to a blue cat (the referent), as she was always the speaker. A man was also present, acting as hearer, and he sometimes was shown looking in the same direction as the woman and sometimes not. Other objects were included for the sake of contrast, such as orange cats that were never the referent. The visual scenes presented to the subjects fit into six categories, divided by their use of relative distance and shared gaze. In the “speaker-associated” (SA) context, the referent blue cat was within the speaker’s arm’s reach. This context would always fit the description of a primary control space, as the speaker would have more action possibilities at her disposal for interacting with the referent than would the hearer. When the referent was placed within arm's length of the hearer (the man), these visual scenes were categorized as a "hearer-associated" (HA) context. A hearer-associated context does not align necessarily with a secondary control space. If the hearer shares the speaker's gaze during this referential act, then indeed we would include the referent object in a secondary control space, as the hearer would have an advantage for physical interaction with the referent if the object is closer to the hearer. A hearer-associated context (i.e., the object is closer to the hearer) can still function as a primary control space if the hearer is attending elsewhere with his gaze, as he would not be able to take advantage of the physical proximity to the referent if it is not in his attention. For the "non-associated" (NA) context, the referent was far away from both the speaker and hearer (i.e. the man and woman were closer to each other than to the cat). Again, the assignment of control space in a non-associated context depends on gaze. If the speaker and hearer are attending in the same direction, then they would together form a primary vantage point and the referent would be seen as distant from them collectively, and the referent would occupy secondary control space far from the speaker and hearer's shared perspective. If the hearer in a non-associated context is looking in a different direction than the referent, then an ambiguity emerges. The speaker may still treat the referent as distant from the location of discourse and assign it a form correlated with a secondary control space (e.g., "that), or the speaker may interpret the referent within her own, primary control space and assign it a correlated form (e.g., "this") because the hearer cannot as readily act on the object without attending to it. Gaze is used thus as a variable: half the pictures show the man (the hearer) looking in the same direction as the speaker ("shared gaze"); in half the pictures, the hearer's gaze is directed in a different direction ("no shared gaze").

The combination of visual and auditory stimuli includes the following. The visual scenes are divided by spatial distance into three contexts (SA, HA, and NA). Another division among the visual scenes is whether or not the hearer shows a shared gaze. The visual scenes were each paired with a verbal phrase, "this one" or "that one," resulting in twelve total possible pairings of visual scene and demonstrative phrase. (Fig. 3.1).

3.2.3. Procedure.

The audio-visual stimuli were presented simultaneously with EEvoke software (ANT Inc., the Netherlands) (Fig. 3.1). The images were displayed on a 1900 ViewSonic LCD monitor at approximately one meter in front of the participant with a 2 ms response time. The subjects were placed in an acoustically and electrically treated room (ETS-Lindgren Acoustic Systems) and were seated in a comfortable chair such that the center of the top of the monitor is level with their eyes. The visual scenes were presented for 2000 ms on the screen with an inter-stimulus interval of 1000 ms. A keyboard (DirectIN-PCB keyboard from Empirisoft Corp.) was placed on the patient's lap for their responses.

Sounds were delivered through bilateral earphones with a presentation level at 60 dB sensation level calibrated per individual on the basis of their hearing threshold (Rao, Zhang, & Miller, 2010). The participants completed the experiment in 1.5 hours, split into five sessions with breaks. The participants were given instructions in both auditory and visual modalities that the scenes would include a man and a woman, and that the woman would always be referring to the blue cat in the scene, despite other objects being present, with one of the following expressions: "this one" or "that one". Patients were requested to press one of two buttons on the computer keyboard to indicate whether the visual scene and demonstrative expression were an expected or unexpected pairing.

The procedure for this experiment differs from other possible procedures that might use the visual stimulus to prime the auditory information. A simultaneous presentation of auditory and visual stimuli has been used in previous experiments, both with the use of ERP and fMRI, as the presentation of both modalities together provides a more natural experience for the subjects (de Gelder, Böcker, Tuomainen, Hensen, & Vroomen, 1999; Ullsperger, Erdmann, Freude, & Dehoff, 2006). The simultaneous audio-visual paradigm was chosen after a pilot priming procedure that conducted for three subjects in addition to those whose data were used in the final analysis and who rated the simultaneous paradigm higher than a sequential presentation. The subjects first underwent a familiarization phase, exposing them for two minutes to the various visual configurations, including the varied spatial distances and shared gaze vs. non-shared gaze conditions in combination with the two verbal stimuli. Following the familiarization task was a test phase consisting of 100 trials of the images and verbal expressions presented in random sequence. The subjects were instructed to judge the semantic congruency of each audio-visual pairing and press the appropriate computer key as soon as possible to signal agreement or disagreement that an image and demonstrative expression form an expected pairing.

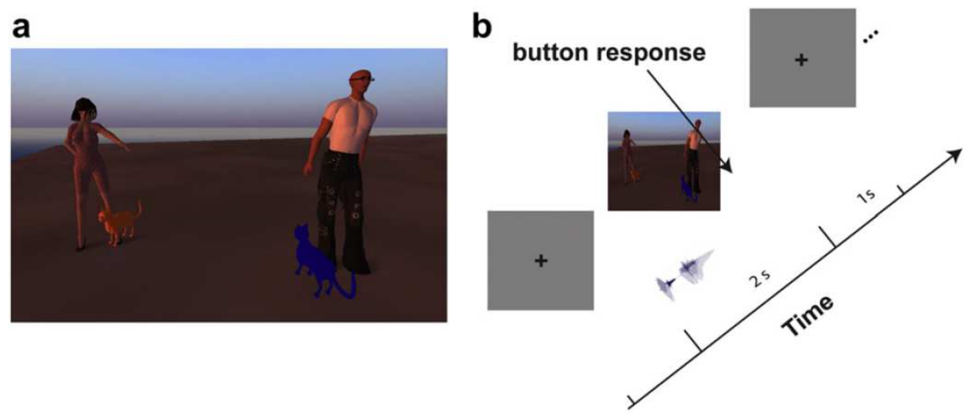

Fig. 3.1. Visual stimulus and presentation cycle. (a) An example of a visual scene presented to the subjects. The female speaker is referring to an object in a Hearer-associated (HA) context without shared gaze (b) Schematic illustration of presentation protocol. The visual scene lasted 2000 ms, and it was simultaneously presented with a female voice referring to the blue cat as either “this one” or “that one.”

3.2.4. EEG recording.

The EEG data were collected with the use of a 64-channel Advanced Neuro Technology system, which had been used in previous ERP studies with auditory stimuli (Rao et al., 2010; Zhang et al., 2011). A 512 Hz sampling rate was used, and the bandpass was between 0.016 and 200 Hz. The AFz site was used as the ground electrode. The EEG cap included shielded wires for 65 Ag/AgCL electrodes, which followed the configuration of the international 10–20 montage system and intermediate locations. The common average of the connected unipolar electrode inputs served as the reference for the amplifier. The impedence for the electrodes was maintained below 5 kOhm. The event markers for the ERPs were time-locked to the onset of the picture and auditory stimulus presented simultaneously.

3.2.5. Behavioral data analysis.

Behavioral data were recorded for each trial of all the subjects. Given the possibility of ambiguity using the control space model, the congruency of a scene-demonstrative pairing was based on the traditional models based on distance. If the referent object was in the speaker-associated space, the congruent response expected was when the demonstrative phrase was “this one” and incongruent when it was “that one.” When the referent object was located in hearer-associated or non-associated space, the congruent response would occur when the demonstrative phrase heard was “that one” and incongruent when the phrase was “this one.” Reponses that conformed with the these congruency parameters would be categorized as a match (hit) or a mismatch (correct rejection). Sensitivity (d-prime) was calculated according to signal detection theory, which treats the formal of an internal representation sensory process of forming an internal representation from a stimulus in order to make a decision (Macmillan & Creelman, 1991). In order to control for false positives and perceptive biases, a comparison was made of d-prime values, rather than percent correct responses. Repeated-measures ANOVA tests were conducted on the d-prime values to evaluate for significance across conditions for response rate. The factors used in the ANOVA tests were spatial context (SA, HA, NA), gaze (share gaze, no shared gaze), and congruency effect (congruent, incongruent). Identical ANOVA tests were used to average reaction time data. When a significant interaction was found, post-hoc two-tailed t-tests were applied to find those subcategories of spatial contexts that showed a significant congruency effect.

3.2.6. ERP waveform analysis.

ERP analysis was conducted with the use of common average reference in ASA 4.7.3 package (Advanced Source Analysis, ANT Corp.). ERP epochs were designated as 900 ms windows with a 100 ms pre-stimulus baseline. When potentials surpassed 50mV in magnitude, these events were rejected offline as artifact. A bandpass filter was applied between 0.5-40 Hz. The minimum average number of accepted trials was 50 for a subject in any of the spatial contexts. Weighted averaging of accepted trials was applied to calculate the grand average. The influence from individual subjects with fewer trials was thus mitigated (Zhang et al., 2011).

Six scalp regions were designated to consider local effects by the following groupings of electrode sites (Fig. 3.2). A right temporal-parietal (RTP) group was formed from the following electrodes: TP7, CP5, CP3, P7, P5, and P3, and the left hemisphere counterparts comprised a left temporal-parietal (LTP) group. A right parietal-occipital (RPO) group was formed from PO7, PO5, PO3, and O1, and a left parietal-occipital (LPO) group was formed from the counterpart sites on the left hemisphere. A centro-parietal midline group (MID) included CPz, CP1, CP2, Pz, P1, and P2. The electrodes POz and Oz comprised a midline parietal-occipital group (MPO). Similar grouping of electrode sites have been used in previous studies (Rao et al., 2010; Zhang et al., 2011).

The area under the ERP waveforms was analyzed for set time intervals, rather than using peak analysis at regional electrode sites. Area analysis was chosen because the ERPs in this study showed late, slow deflections with clearly defined peaks for most subjects' data. A point-to-point t-test was used for ERPs in the congruent and incongruent conditions to determine the time windows to investigate. The time windows were defined on the basis of time intervals that the Pz electrode showed significant differences between the congruent and incongruent conditions (e.g., Kutas & Hillyard, 1980). The time interval would be said to show a congruency effect if a significant difference ($p < 0.05$) was observed for minimally 40ms (Zhang et al., 2011).

Repeated-measures ANOVA tests were conducted in Systat (Version 12) to assess the main factors spatial context, congruency, scalp region, and laterality. Laterality was tested using right hemisphere sites (RFC, RTP, RPO) versus left hemisphere sites (LFC, LTP, LPO). Greenhouse-Geisser corrections were applied wherever appropriate. A modified two-tailed point-to-point t-test procedure that is similar to Bonferroni correction was applied to the regional electrode groupings to evaluate the temporal nature of differences between the congruent and incongruent conditions (Guthrie & Buchwald, 1991; Zhang et al., 2011).

3.2.7. Global field power analysis.

The Global Field Power (GFP) shows an overall effect across the scalp independently of electrode location. This measure was calculated for the experimental conditions of all the participants' datasets. The GFP describes the distribution of potentials for all 64 electrodes over the scalp in terms of their standard deviation at every sampling point during the ERP epoch, one is able to (Hamburger & Burgt, 1991; Lehmann & Skrandies, 1980). The sampling points for the congruent and incongruent conditions were converted into z scores relative to the distribution of baseline GFP 100 ms pre-stimulus (Rao et al., 2010). A congruency effect was identified for latency intervals where a significant difference between the congruent and incongruent conditions persisted for at least 40 ms.

3.2.8. sLORETA analysis.

Standardized low-resolution brain electromagnetic tomography (sLORETA) was used for source estimation of the ERP data with use of the ASA software (Version 4.7.3). LORETA uses an inverse solution to localize distributed sources of activity (Pascual-Marqui, Michel, & Lehmann, 1994). An improved algorithm is used for the present data that standardizes the signal-to-noise ratio and therefore minimizes the effect of inherent noise in the EEG signal in source localization (Pascual-Marqui, Esslen, Kochi, & Lehmann, 2002; Sekihara, Sahani, & Nagarajan, 2005).

The sLORETA analysis followed the following steps:

1. sLORETA analysis was applied for the averaged ERP data for every individual and for each condition.

2. Standard positions for electrodes were taken from a healthy adult subject's MRI structure to produce a volume conductor head model. The Boundary Element Model method was used to form a three-layer head model with consideration for the geometric and electrical properties of head anatomy, allowing one to work with a realistic head model for source localization (Zhang et al., 2009, 2011; Zhang, Kuhl, Imada, Kotani, & Tohkura, 2005).

3. Source spaces were modeled as surfaces on which dipoles were equidistantly distributed and interconnected with triangles to display dipole magnitude by color and contour. Three orthogonal dipoles were placed at one location to make dipole orientation a free parameter as well. The dipoles were combined to one at pre-specified locations during the final amplitude computation for statistical analysis. The grid spacing for the

sLORETA metric was set at 20 mm (Zhang et al., 2009). Tikhonov regularization, computed by the Generalized Cross Validation method, was used to smooth out the source current distribution, with balanced sensitivity to superficial and deep current (Bortel & Sovka, 2007; Tychonoff & Arsenin, 1977).

4. sLORETA amplitude data for all dipoles and the relevant spatial coordinates were exported in ASCII format and analyzed for regions of interest (ROI). To define the ROIs, coordinates were transformed to the standard volume of Talaraich coordinates (Talairach & Tournoux, 1988). Spatial coordinates were then converted into Brodmann Area with the use of the Talaraich Client (www.talairach.org) (Lancaster et al., 2000). The ROIs for the present project, given it being a language task, include the parietal lobe (Brodmann Areas 5,7), left middle and superior temporal gyri (BA 21,22), and the parahippocampal gyrus (BA 36). The sLORETA amplitude for these ROIs were each calculated as the sum of the dipole activities for that region.

5. The sLORETA data for the ROIs were then analyzed with repeated-measures ANOVA, after which post-hoc analysis between the congruent and incongruent conditions at each ROI and for each spatial context was performed with Bonferroni-corrected two-tailed t-tests.

3.3. Results.

3.3.1. Behavioral data.

Repeated measures ANOVA test for behavioral conformity data did not show significant effects for three main factors: congruency, spatial context (SA, HA, NA), and gaze (shared vs. non-shared). A high level of consistency (above 95% acceptance) was observed for each type of congruent pairings (Table 3.1). The level of consistency for incongruent pairings was over 90% rejection. Corresponding d-prime scores were above 2. The reaction time data (Table 1) showed significant main effects for congruency (congruent vs. incongruent) [$F(1,13) = 34$, $p < 0.0001$] and spatial configuration (SA, HA, and NA) [$F(2,26) = 11.3$, $p < 0.001$]. A significant interaction between spatial context and the shared gaze condition was found as well, [$F(2,26) = 7.8$, $p < 0.01$]. Post-hoc t-tests revealed a role for shared gaze for the congruency effect, such that only in the shared gaze condition were reaction times significantly longer for the incongruent pairings than the congruent pairings for each of the spatial contexts (SA, HA, NA), $p < 0.01$. No significant differences were observed in reaction times between congruent and incongruent pairings without shared gaze between the speaker and hearer.

		Shared gaze			No shared gaze		
		SA	HA	NA	SA	HA	NA
RT (ms)	Congruent	897 (s.d. 178)	891 (s.d. 173)	931 (s.d. 228)	897 (s.d. 213)	994 (s.d. 227)	1064 (s.d. 250)
	Incongruent	1005 (s.d. 193)	985 (s.d. 205)	1064 (s.d. 250)	984 (s.d. 205)	1039 (s.d. 221)	1031 (s.d. 197)
Conformity (%)	Congruent	97	96	96	97	95	96
	Incongruent	91	88	96	95	90	93
d′ values	Congruent	2.3	2.1	2.7	2.8	2.2	2.6
	Incongruent	1.9	2.2	2.8	2.6	2.3	2.6

Table 3.1. Behavioral data of English demonstratives per gaze and congruency conditions. Data include mean reaction times (TRT) with standard deviation (s.d.), mean percent response conformity with expected demonstrative form, and mean d0 values for SA (speaker-associated), HA (hearer-associated), and NA (non-associated) context with and without shared gaze in congruent and incongruent conditions. Congruency effect was observed in RT data for SA, HA, NA contexts in the shared gaze conditions.

3.3.2. ERP waveform data.

The ERP waveforms data were tested in repeated-measures ANOVA with the following main factors: congruency, spatial context, gaze, and laterality (left vs. right electrode sites) (Fig. 3.2). Significant effects for spatial context [$F(2,26) = 13.3$, $p < 0.001$], gaze

[F(1,13) = 8.4, $p < 0.05$], and laterality [F(1,13) = 6.1, $p < 0.05$] during the 525-725ms time window. A significant interaction between congruency and spatial context, [F(2,26) = 7.7, $p < 0.01$], suggests that congruency decisions depend on spatial context. ANOVA tests for the earlier intervals (100–200 ms, 200–300 ms, 300–500 ms) showed no significant effect for congruency. ANOVA tests were then tested for the three spatial contexts separately for a congruency effect as a function of electrode location and gaze. In the HA context with shared gaze, a significant congruent effect was found, [F(1,13) = 4.8, $p < 0.05$], as well as a significant interaction of congruency and electrode region [F(8,104) = 4.5, $p < 0.001$]. Follow-up point-to-point t-tests showed a significant congruency effect at centro-parietal electrode sites, including MID [$p < 0.01$], MPO [$p < 0.01$], LPO [$p < 0.001$], RPO [$p < 0.001$], LTP [$p < 0.01$], and RTP [$p < 0.05$] (Fig. 3.2). In contrast with the behavioral data, the SA and NA contexts did not reveal a significant congruency effect in the ANOVA tests with shared gaze.

3.3.3. Global field power data.

The data from the global field power showed a significant congruency effect at different time intervals for different spatial contexts (Fig. 3.2). The SA contexts, similarly to the ANOVA analysis, did not show significant congruent effects with or without shared gaze in terms of converted z scores for the GFP data. In the HA context with shared gaze, a significant congruency effect was observed during the windows 200-500 ms, 575-615 ms, and 730-800 ms. For the NA context with shared gaze, a significant congruency effect was found during the GFP data during the 280-760 ms time window.

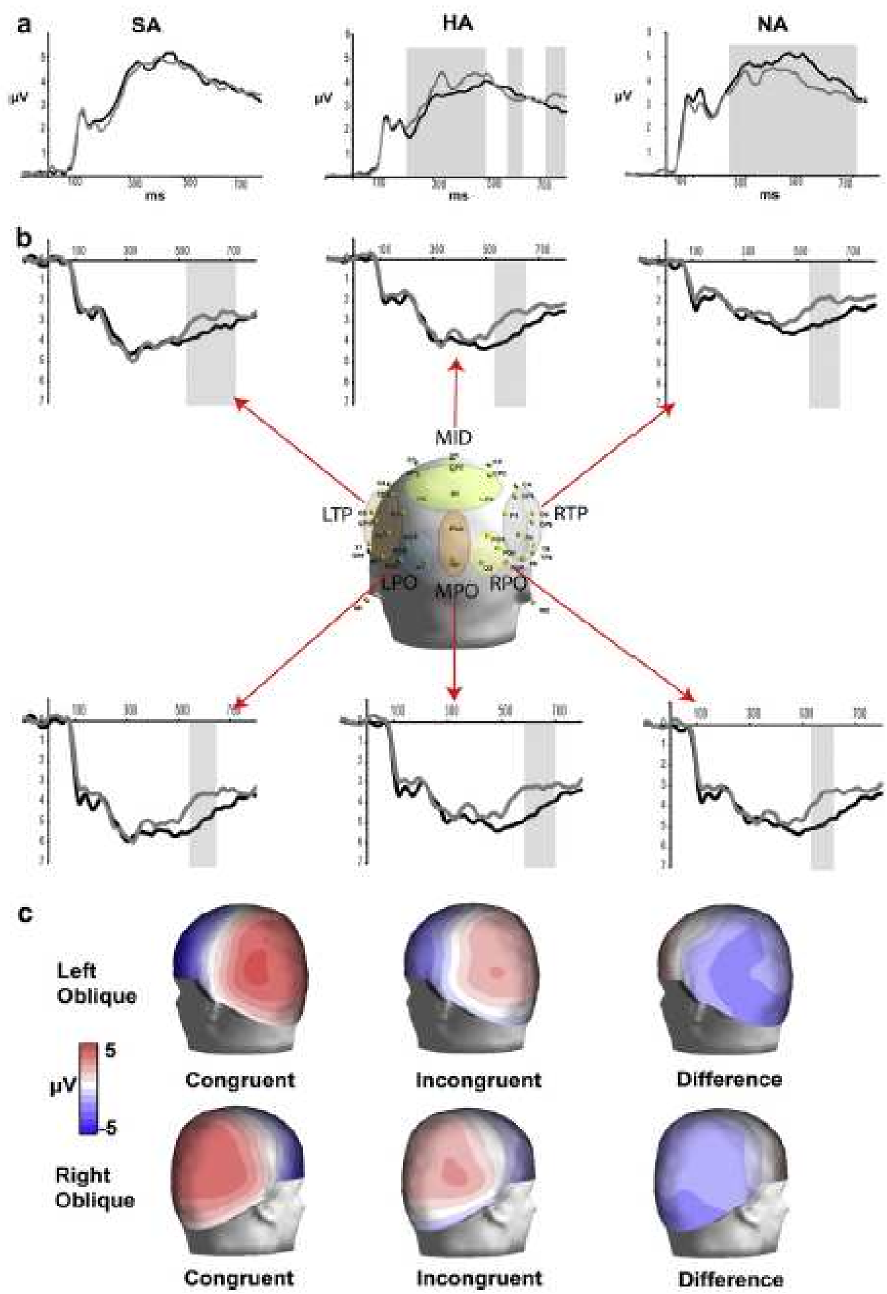

Fig. 3.2. Waveform analysis of gesture condition. (a) Global field power data, left to right, for the speaker-associated context, hearer-associated context, and non-associated context, all with shared gaze. The black line is the congruent condition and the gray line is the incongruent condition. Shaded areas show time intervals of at least 80 ms at the significance level of 0.01 (b) Grand average ERP waveform amplitude (mV) from six electrode groups during the hearer-associated spatial context with shared gaze with head model to show ROIs. The black line represents the congruent condition, and the gray line the incongruent condition. The grouping sites on the scalp are marked in the head model. Shaded areas indicate time intervals of minimally 40 ms where the point-to-point t-test with correction for multiple correction was significant [$p < 0.05$] (c) Scalp distribution maps for the HA context at 600 ms. Included are the congruent and incongruent scalp distribution maps, as well as the difference (incongruent – congruent).

3.3.4. sLORETA data.

The sLORETA analysis revealed three main regions of activity for the congruency effect: the left middle and superior temporal gyri (BA 21, 22), the parahippocampal gyri (BA 36), and a broad area of activation at posterior regions, including the superior parietal lobule (BA 5, 7), the primary motor cortex (BA 4), and the cingulate gyrus (BA 31) (Table 3.2). Significant effects for sLORETA amplitude results were found in the 525-725 ms interval for the main factor of spatial context, [$F(2,26) = 7.7$, $p < 0.01$], as well as the interaction of congruency and brain region, [F(2,26) ¼ 4.1, $p < 0.05$]. A significant congruency effect was found by post-hoc t-tests in three cortical areas for the HA spatial context with shared gaze: the left temporal gyri [$p < 0.05$], the left superior parietal lobule [$p < 0.001$], and the right superior temporal lobule [$p < 0.01$] (Fig. 3.3). Parietal activation was therefore bilateral. No significant congruency effect was gathered from the SA and NA spatial contexts during the 525-725 ms window irrespective of gaze.

Brain region	BA	Coordinates	Congruent sLORETA amplitude	Incongruent sLORETA amplitude	*p*-value
Parietal lobe, Pre-central gyrus, Cingulate gyrus	4, 5, 7, 31	10, −38, 56	15.8 nA	5.4 nA	***
Parahippocampal gyrus	36	39, −21, −16	17.9 nA	25.7 nA	n.s.
Left temporal gyri	21, 22	−60, 2, −10	35.2 nA	25.4 nA	*

Table 3.2. sLORETA data per congruity condition in the hearer-associated context for three regions of interest. Spatial locations with Brodmann area (BA) and coordinates were determined in the Talairach space. ***stands for $p < 0.001$, *for $p < 0.05$, n.s. for not significant.

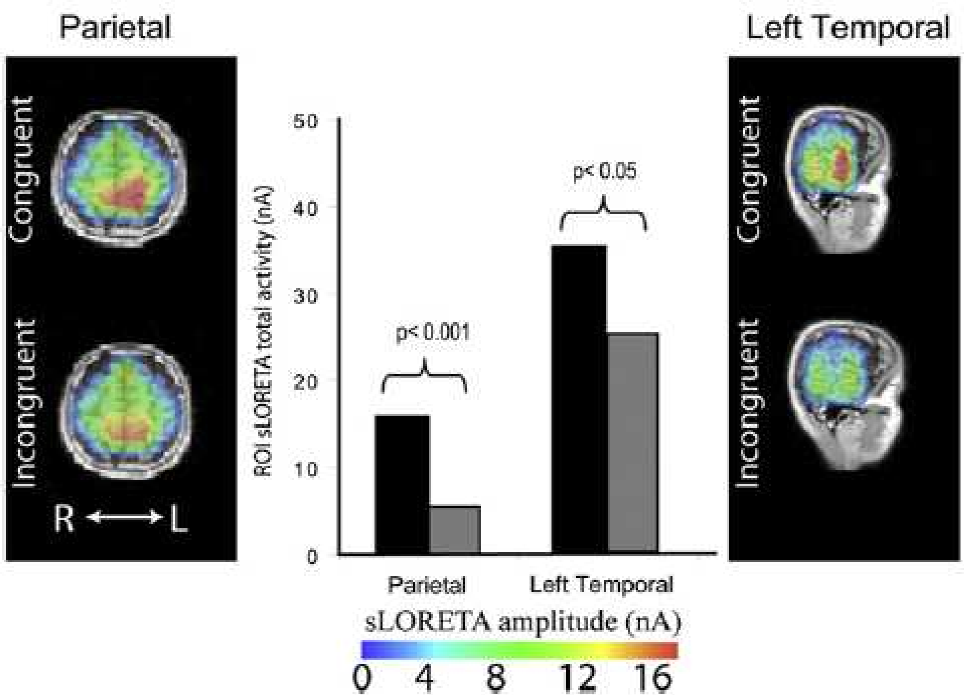

Fig. 3.3. Grand mean sLORETA results for hearer-associated space with shared gaze. Significant differences were observed between the congruent and incongruent conditions at a broad posterior region and at the left middle and superior temporal gyri. Stronger activation for the congruent condition was found at the left temporal cortex and the parietal lobe bilaterally.

3.4. Discussion.

As an experiment to pursue difficult questions regarding the function of demonstrative use, we have some evidence to approach the questions freshly. No single experiment will prove or even substantiate a theoretical framework such as that of control space, but the experiment is able to test individual hypotheses. As discussed in Chapter 1, the control space model would predict for the acceptability of a demonstrative form to depend on relative distance, so long as the interlocutors have shared gaze. A related hypothesis predicted from the control space concept is that relative gaze should serve less in demonstrative form selection if no shared gaze is maintained between speaker and audience. Another hypothesis, which is not specific to any model of demonstrative use, is that incongruent pairings of demonstrative form and visual context would elicit an N400. The finding of an N400 would further support the interpretation of this component as an index of unexpected data for a given context.

3.4.1. Dependency of relative distance and gaze for demonstrative use.

There are reasons in favor and against the traditional account for spatial demonstratives in English that describes the use of “this” and “that” as connoting near and far space, respectively (Halliday & Hasan, 1979; Lyons, 1977; Quirk, 1979). The role that gaze plays in the acceptance of these forms gives reason for caution with a distance model. The control space account does, however, expect for distance to play a role in form selection, and the data presented above support this claim. The behavioral accuracy data in the present study were consistent with the view that demonstrative choice makes use of physical space as a division, particularly between near contexts (SA) and that for far contexts (HA, NA).

Other evidence has supported the distance-based distinction in demonstratives, such as the psycholinguistic work by Coventry et al. (2008). The authors argue in fact for an anatomical basis for the spatial correlates of demonstrative forms, based on how the human visual system processed near and far space differently. This argument, however, does not clear take into account the importance of shared gaze. The authors themselves do not explain how an anatomical account of demonstrative use would adequately describe their own observation that pointing tools could mitigate the role of perceptual distance.

The data from the present study show that only with shared gaze did the response times between congruent and incongruent pairings differ significantly. These results coincide with Diessel's (2006) claim that English speakers consider joint attention in the selection of demonstrative forms. The present data suggest that neither joint attention nor spatial distance alone can explain English demonstrative use, as the role for attention cannot in itself serve as a complete replacement for relative distance. The response time data from the present project do show that distance from the speaker affected demonstrative form acceptance when shared gaze was present. Interestingly, no significant differences were found between "this one" and "that one" when the hearer was looking elsewhere. The role of joint attention in English demonstrative selection would not be unique to that language. In Turkish, a language with three demonstratives, "bu" and "o" depend on relative distance, yet a third form, "su," relates primarily to the hearer's attention (Küntay & Özyürek, 2006).

Distance from the speaker not only cannot serve as an adequate predictor of demonstrative form because of the mitigating role for gaze but also because space from the hearer is relevant. The ERP results reveal a significant congruency effect in the latency window of 525-725 ms during the hearer-associated (HA) context with shared gaze. In other words, participants seem to be more sensitive in their acceptance of the distal form, "that one," when used close to a hearer than when used for referents distant from the conversation.

The congruency effect in the ERP data confirmed the hypothesis that an N400-like component would be elicited to unexpected pairings of demonstrative and visual context. The ERP was observed over the posterior scalp with maximum amplitude approximately at 600 ms post-stimulus. The later latency of the ERP is understood to result from the complexity of the task: interpreting verbal information in conjunction with a visual scene that includes such important landmarks as a speaker, hearer (with gaze variation), and referent, among various distracters. The centro-parietal scalp potential distribution of the present ERP finding and its finding in response to varied spatial contexts resembles the Gesture N450 effect reported by Wu and Coulson (2005). The longer latency of the present data may result from the subtlety of the decision to accept a demonstrative for spatial reference, a form of expression that may be vaguer than verbal-only descriptions.

The complexity of the task in judging spatial use of demonstratives may also inform the differences in results from the behavioral data, the ERP waveform, GFP, and sLORETA analyses. The reaction time data showed a congruency effect for all spatial contexts with shared gaze. The ERP data, however, showed a significant congruency effect for the HA context. The SA and NA contexts, according to the grand average (Fig. 3.2), have the appearance of an effect that was not borne out statistically, a pattern that may point to variability among subjects' brain responses. Minor discrepancies in results among the ERP waveform, GFP, and sLORETA may also be due to differences in technical procedure. For instance, the GFP is calculated on the basis of ERP data from all electrodes, whereas the sLORETA makes use of a source localization algorithm that is sensitive to the signal-to-noise ratio of the ERP data. Despite differences in results by type of analysis, the ERP waveform, GFP, and sLORETA data all showed a late negative congruency effect in the HA context with shared gaze that would be consistent with previous N400 reports.

Notably, the sLORETA analysis indicated a cortical network for the late N400 effect found in this study, namely, the parahippocampul gyrus, the left middle and superior temporal gyri, and a broad region of parietal lobe bilaterally. Consistently with the waveform analysis, the HA context with shared gaze was the only one to show a significant congruency effect per sLORETA. The active regions in the present study have been found to associate with processes of agency, attention, the distinction between self and others, and perception-action coupling procedures involved in social behavior (Decety & Lamm, 2007; Sommerville & Decety, 2006). The parietal cortex has also been found to be active during semantic integration, according to fMRI data (Chou et al., 2009; Grossman et al., 2003). Although previous work has found the inferior parietal lobule to be involved especially in semantic processing, the present study found activation more properly in the superior parietal lobule (Chou et al., 2009), a site known to process spatial relationships between our hands and hand-held objects (Naito et al., 2008). The superior parietal lobule has further more been implicated during reaching actions among primates (Caminiti, Ferraina, & Johnson, 1996) and during observed and imagined reaching among human subjects (Filimon, Nelson, Hagler, & Serenoa, 2007).

The other areas of activation in the present work can be easily understood for the task. The left temporal cortex, for instance, is a known site for semantic processing (Boddaert et al., 2004; Simos et al., 1999). The parahippocampal gyrus has been found to be associated with the negotiation of physical places while performing a task (Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998). Epstein and Kanwisher found this area to be engaged by viewing scenes but not in processing faces, objects, or faces. The need to calculate the spatial positions among figures in the visual contexts of this experiment may explain activation of the parahippocampal gyrus.

The study discussed here offers new information regarding long-standing debates regarding demonstrative use. The traditional view of English demonstratives considers the forms “this” and “that” to differ according to distance from speaker, whereas more recent accounts have employed concepts of attention and focus to describe the functions of demonstratives separately. Data from the present experiment are that spatial distance alone does not seem to sufficiently define demonstrative terms in English. The behavioral reaction time data did find a congruency effect, based on distance, but this held true only for visual contexts where the speaker and hearer shared gaze. The ERP data further revealed a late negative deflection (525-725 ms post-stimulus), assumed here to be a variant of the N400 component, to index the unexpected use of demonstratives in the hearer-associated context (i.e. when the referent was within reach of the hearer) with shared gaze only. The more special contexts in which demonstrative use was processed anomalously, that is, with shared gaze and particularly when the referent was located near the hearer, suggest another explanation for demonstrative forms. If demonstrative use is driven more by how discourse participants analyze their control over space, the present data would support this account, as speakers and hearers would be sensitive to how far they are positioned physically from an object in order to evaluate possible actions toward the object, but this physical distance would matter less if the speaker and hearer do not share gaze and thus neither a shared sense of readiness to interact with an object. The observation that participants were sensitive to the distance of an object not only to the speaker but also to the hearer further suggests a description of demonstrative forms based on their relative ability for future interaction with a referent.

Chapter 4:

The role of co-speech gesture: An ERP study of English and Japanese demonstratives

4.1. Introduction.

In Chapter 3, we saw how the ERP technique was helpful to investigate factors that influence the acceptability of demonstrative forms, including spatial distance of the referent to both speaker and hearer, as well as their gaze. A second electrophysiological experiment is discussed in this chapter, which is a revision of a manuscript submitted for publication (Stevens & Zhang, 2014), that considers additional basic questions regarding an expected demonstration. The concept of a control space, some restriction on physical or abstract space in which different agents can evaluate their possibilities for action, was used to explain the English speakers' sensitivity to spatial distance and gaze toward an object in their acceptance of demonstrative forms. A useful extension of this concept is to employ data from other languages. Japanese offers an interesting system to compare with English, such that Japanese has three demonstrative morphemes, こ("ko"), そ("so"), and あ("a"). An initial question would be whether a different language with divisions of form unlike English would also reaction time differences and an N400 effect when forms did not match visual contexts according to spatial distance. A second question that is basic to demonstrative use that has been discussed extensively above is gesture, and so it would be worthwhile to observe the effect of withholding a pointing gesture while using a demonstrative for spatial reference. The use of a pointing gesture to help establish a reference would be consistent with the concept of a control space, as a control space will depend on such tools to restrict its dimensions as a gesture.

The most long-standing explanations for English and Japanese demonstrative use make use of spatial distance to distinguish forms. English demonstratives have been explained in terms of distance from the speaker, such that "this" serves to refer to objects that are relatively near the speaker and "that" to pick out things that are not near the speaker. The traditional explanation of the Japanese demonstrative system is that "ko," "so," and "a" pick out an object that is close to the speaker, close to the hearer, and far from both, respectively, if the speaker and hearer do not have a shared perspective (Aoyama, 1995; Matsushita, 1901; Sakuma, 1936). The Japanese system differs from English by formally encoding the hearer's position.

The difference between the English and Japanese demonstrative systems suggests the possible influence of a language on categorization of space. The role that an individual language plays on the processing of the physical environment has been investigated for various categories, such as color. Winawer et al. (2007) experimentally confirmed the observation that English and Russian speakers discriminate colors differently for shades of the English blue by asking participants to choose which two of three color squares were identical. Native Russian speakers were able to discriminate between shades of blue corresponding to "siniy" and "goluboy" with a quicker reaction time than native English speakers. Notably, English speakers do divide light and dark blue at a similar point on the color spectrum to Russian "siniy" and "goluboy", but Russian speakers discriminate obligatorily for reasons of communication.

Spatial relationships have been argued to show the effect of linguistic categorization (Bowerman, 2007; Clark, 1973; Landau & Jackendoff, 1993; Levinson, 2003; Lucy, 1992). Due to the abstractness of English prepositions, in which few physical constraints are placed on objects, Landau and Jackendoff (1993) take the distinctions to reflect universal distinctions of geometry. One argument for children having innate spatial reasoning reflects on the observation that children try to express spatial relationships before having the relevant lexicon, e.g., “towel bed” for a towel on a bed (Bloom, 1970). Similarly, children spontaneously overextend and underextend spatial concepts, a finding that argues against language strictly determining categorical division, but have been observed to do so in language-specific ways. English-speaking children, for instance, use ‘open’ to describe pulling apart two Frisbees (Bowerman, 1978). To investigate, Bowerman (1996, 2007) takes the case of spatial words acquired early by children, e.g., ‘in’, ‘on’, which presuppose a relationship in which a figure object interacts with a ground object, the latter classifiable as container or surface. Bowerman (1996) reports that children overgeneralize concepts in accordance with patterns in their language. For instance, Dutch children use the word “uit” to describe removal from a surface, based on analogy to adult use to take off clothing. Adult speakers of French and English have shown differences in analyzing objects in figure-ground contrasts (Hickmann, 2007). French speakers were less able than English speakers to describe a door handle’s location in relation to the door or to describe the location of a crack on a cup. In light of such language-dependent performance, Bowerman (2007) claims that spatial situations contain multiple properties that can serve to relate objects, and languages can focus on such properties idiosyncratically.

Effects of language experience have also been observed in the context of directions. In English, directions are given often on the basis of a point of origin. For example, the expression “on the left” places a destination relative to the location of the utterance. Besides a relative reference frame, it is possible to describe spatial relationships in intrinsic terms of an object (e.g., ‘in front of the TV’) or by an absolute reference frame (e.g., “north of the lake”) (Levinson, 1996). Languages differ by whether they use mostly or exclusively one reference frame (Levinson, 2003). Speakers of languages with absolute reference frames would express a location in such terms as “the fork is to the south of your foot.” The absolute directions in such languages are not necessarily precise. The Tzeltal of Mexico use a word translated as ‘uphill’ to denote 345° N. Guugu Yimithirr of Australia uses a local ‘north’ to denote 17° N (Majid, Bowerman, Kita, Haun, & Levinson, 2004).

In light of the above research into the relative effects of language on one’s interaction with the physical environment, it would be helpful to investigate how different demonstrative systems influence the interpretation of space. One means to consider language-specific effects of demonstratives is to look at the use of accompanying tools, and demonstratives are ubiquitously expressed with a pointing gesture. The concomitant use of gesture with speech may compensate for information not expressed in speech and highlight congruent semantic information (Holle & Gunter, 2007; Wu & Coulson, 2005, 2007). The crossmodal compensation hypothesis or Mutually Adaptive Modalities hypothesis (De Ruiter, 2006) views language and gesture functioning together in a compensatory relationship to ensure communication. As an example that gesture spontaneously complements language, a letter to the editor of the New York Times by retired radiologist Steven Sitzman (2013) shares about his wife’s temporary paralysis from Guillain-Barré Syndrome, such that after her recovery he was struck by her immediate use of her hands when she spoke.

There remains much to understand regarding how speech and gesture function across languages. Languages may require degrees of gesticulation to express a concept, a possibility consistent with the Mutually Adaptive Modalities hypothesis (De Ruiter, 2006; Melinger & Levelt, 2005), which interprets gestures will complement speech more in settings where speech is less effective, such as in a setting with more ambient noise, and conversely that gestures are used less when speech is more effective, such as over the telephone. Similarly, it has been found that speakers gesture less when their descriptions of an object are more detailed (Melinger & Levelt, 2005). When considering the Mutually Adaptive Modalities Hypothesis across languages, the informative value of the pointing gesture might be predicted to differ. Specifically, Japanese speakers may rely less on a pointing gesture to evaluate reference than English speakers, because they have more precise lexical means to isolate the intended object.

Additional theoretical proposals relating visual, verbal, and gestural representations are available. The Lexical Semantics Hypothesis (LSH) claims that gesture generation stems from the semantics of the lexicon to convey the desired message (Butterworth & Hadar, 1989; Schegloff, 1984). The LSH predicts that gestures precisely correspond to the specific meaning expressed by the verbal expressions: just as the semantics of a word can drive its syntactic position, a demonstrative's spatial feature will generate a gesture to assist in the localization of the referent. Since the gesture results from the semantic structure of demonstratives according to the LSH, the absence of a pointing gesture with the use of a spatial demonstrative may combine to produce a semantic anomaly. By contrast, the Free Imagery Hypothesis assumes that gestures are based on pre-linguistic and non-propositional imagery representations and thus would be unaffected by the language production processes that convert the imagistic representation into propositional content (Krauss, Chen, & Chawla, 1996; Krauss, Chen, & Gottesman, 2000).

The three hypotheses above regarding speech and gesture co-occurrence describe spontaneous production, yet perception of demonstrative use can offer information regarding the production of gesture if the expectations of the observer are not met. The hypotheses are also not concerned with the electrophysiological responses generated in a perceptual paradigm using expected and unexpected pairings of gesture and language, but their claims frame questions of the cognitive processes of demonstrative use in different languages and with gesture. Gestures both impact a language user's production and inform comprehension, and gesture handling requires cognitive processes to mediate the different knowledge levels including motor skills, the choice of linguistic expressions, and social intentions (Kelly, Manning, & Rodak, 2008; Kendon, 2004; McNeill, 1992).

As with the project described in Chapter 3, the cross-linguistic study discussed presently also sought to investigate the brain mechanisms behind coordinating speech, space, and gesture by comparing event-related potentials (ERPs) time-locked to the onset of an audio demonstrative expression, which was simultaneously presented with an artificial visual scene including a speaker, a hearer, and the referent object. The combination of the audio demonstrative with the visual scene was classified as either congruent or incongruent based on the expectation from traditional accounts of demonstratives in English and Japanese indicating relative distance (Aoyama, 1995; Halliday & Hasan, 1979; Lyons, 1977; Matsushita, 1901; Quirk, 1979; Sakuma, 1936). In addition to the spatial position of the object of reference to the interlocutors, another parameter of interest was the presence or absence of a pointing gesture in the visual scenes. The basic premise in adopting the perceptual experimental paradigm and manipulating gestural presence is in line with the cognitive computational model for embodied gesture processing (Sadehipour & Kopp, 2011). Production models of gesture and language can be addressed through their perceptual interpolations, as the product of the cognitive processes involved is observable behavior that can be evaluated by an interlocutor (and language learner) and serves as the data driving language production.

In the previous study (Stevens & Zhang, 2013), the mismatch effect between lexical form and visual context produced a late N400 effect (525-725 ms) when a demonstrative was presented in an incongruent context with shared gaze within arm's reach of the hearer, and the associated scalp topography and source localization corresponded with the patterns uncovered from other N400 studies of semantic or contextual anomaly (Kutas & Hillyard, 1980, 1984). Several other ERP studies have also addressed congruent and incongruent gesture use with verbal expressions (Cornejo et al., 2009; Gredebäck, Melinder, & Daum, 2010; Holle & Gunter, 2007; Kelly et al., 2004, 2007; Özyürek et al., 2007; Wu & Coulson, 2005, 2007). However, none of the existing ERP studies addressed brain mechanisms governing the use of spatial demonstratives in different languages.

This investigation had three specific objectives. (1) To investigate whether an unexpected spatial context for a demonstrative will elicit the N400-like congruency effect (i.e., significant differences between the congruent and incongruent trials) in both English and Japanese subject groups for the trials with a pointing gesture. (2) To examine how the behavioral and ERP responses are affected by the absence of a pointing gesture. (3) To compare the findings in English and Japanese for the waveform morphology, latency, and topographical distribution of the N400 or P600 responses. The first hypothesis is that participants from both language groups recognize unexpected demonstrative use partly on the basis of relative distance, and therefore a congruency effect should be observed in both English and Japanese behavioral reaction time data (i.e., longer response time for incongruent trials than for congruent trials) as well as by a N400-like ERP component. An N400-like component was observed in the previous study (Stevens & Zhang, 2013) but has been elicited elsewhere by both linguistic and non-linguistic stimuli when participants are exposed to surprising contexts, which may depend on semantic or pragmatic knowledge (Hagoort et al., 2004; Van Berkum, Zwitserlood, Hagoort, & Brown, 2003; West & Holcomb, 2002; Wu & Coulson, 2005). A second hypothesis was that speakers of one language would be more sensitive to a referential gesture than those of another language if there are fewer lexical resources by which to describe spatial relations. In particular, English speakers' behavioral and brain responses would be more affected by the absence of a pointing gesture than among Japanese speakers, as English has fewer demonstrative forms and may therefore depend more on extra-linguistic information. This hypothesis, if confirmed, would be in line with the Mutually Adaptive Modalities Hypothesis, which considers speech and gesture to supplement each other for the benefit of efficient communication. This hypothesis would be also consistent with the Lexical Semantics Hypothesis, as an indexical gesture would be understood to be generated from the semantics of the associated demonstrative expression. In contrast, if a pointing gesture or its absence did not significantly modulate brain responses in either subject group, the data would be consistent with the Free Imagery Hypothesis, as gesture functioned as a pre-linguistic symbol.

Although the prediction preceding the project was that incongruent audiovisual pairings would produce an N400 effect, another possible ERP effect to consider for the present match/mismatch experiment is a P600, a broad, positive voltage change peaking at approximately 600 ms post-stimulus elicited not only by morphosyntactic anomalies (Hagoort et al., 1993; Kaan, Harris, Gibson, & Holcomb, 2000; Osterhout, Holcomb, & Swinney, 1994), but also by special cases of semantic violations including picture-expression mismatch (Kim & Osterhout, 2005; Kuperberg et al., 2003; Nieuwland & Van Berkum, 2005; Vissers, Kolk, van de Meerendonk, & Chwilla, 2008). Given the formulaic co-occurrence of speech and gesture in a demonstration, the use of a demonstrative form for spatial reference without the accompanying gesture may constitute a violation of the embodied grammar for gesticulation and thereby elicit a P600. Alternatively, according to the online Monitoring Theory (Vissers et al., 2008), the P600 reflects not just syntactic or structural integration but a more general monitoring mechanism to track potential conflicts or processing errors regarding the expected representation and input information. According to this more general interpretation of the P600, the congruency effect in the present study could result in a late positive effect to index the participants' recognition of a mismatch between a linguistic expression and the context for which it forms an expected pattern.

4.2. Methods.

4.2.1. Participants.

The study followed the informed consent protocol approved by the Institutional Review Board. Twelve adult native English and twelve adult native Japanese speakers were recruited via advertisement. The English speakers (6 males and 6 females) had all had some formal exposure to other language instruction, and the Japanese speakers (5 males and 7 females) were all born in Japan and were conversant in English. All participants were university undergraduate students or graduates with the average age at 29. The Edinburgh handedness inventory index score, averaged across subjects, was +0.85 (right-handed) (Oldfield, 1971). There were significant differences in age, handedness scores between the two subject groups. All subjects reported to have normal or correct normal vision without color blindness and had normal hearing. No subjects had any speech or language disorders. They were paid $25 each for their time. Data from four participants (two English-speaking subjects and two Japanese-speaking subjects) were excluded due to excessive blinking and other muscular artifacts.

4.2.2. Stimuli.

As detailed regarding the experiment described in Chapter 3, the auditory stimuli were similarly synthesized using AT&T text-to-speech software and then digitally edited with Sound Forge 9 (Sony Corp.). The root mean square intensity levels were normalized for the auditory samples. The visual stimuli were made from the Second Life software, which provided a virtual world platform with 3D images (http://secondlife.com). The images contained varied spatial configurations of the speaker, the hearer, and the referent object: a flamingo, table, chair, or lamp (see an example in Fig. 4.1), in which the object was either near the speaker, near the hearer, or farther than both than they were to each other. Gesture use was also varied, such that in half the image, the speaker (the male figure in the images) was pointing to the referent, and in the other half the speaker is not gesturing. The visual scenes in the audiovisual combinations were identical for the two subject groups.

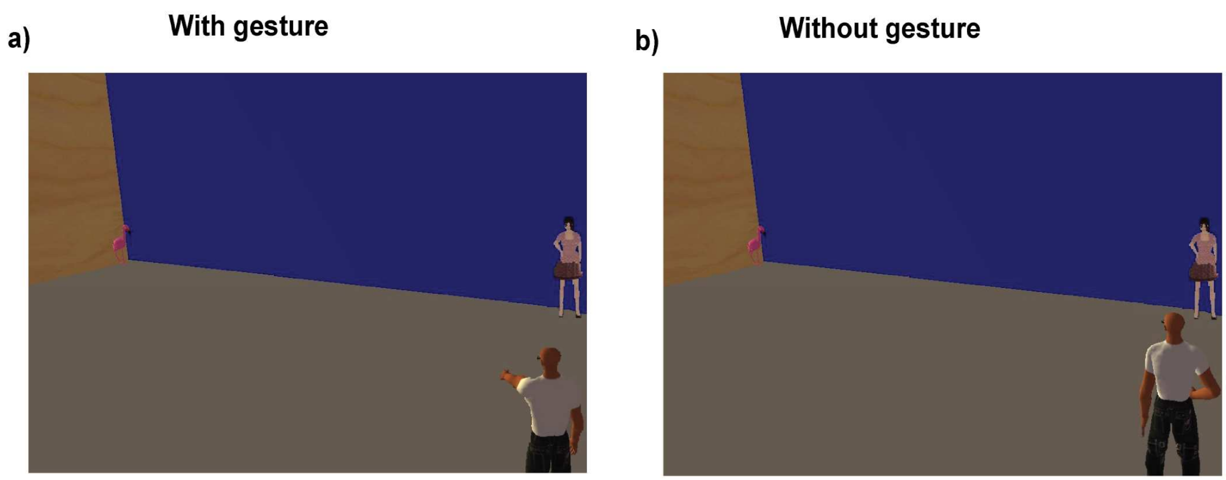

Fig. 4.1. Visual stimuli, with gesture and without gesture. (a) An example of a visual scene for the gesture trials. The male speaker is pointing at an object that is distant to both the speaker and hearer. (b) An example of a visual scene for the no-gesture trials. The speaker refers to the distant object without an accompanying pointing gesture.

4.2.3. Procedure.

The audio and visual stimuli were simultaneously presented in each trial to test the congruency effect (Fig. 4.2) (Stevens & Zhang, 2013). The audio-visual stimuli were presented with the EEvoke® software (ANT Inc., the Netherlands). The pictures were displayed for a duration of 2250 ms on a 19” ViewSonic LCD monitor (1280×1024 at 60 Hz) located approximately one meter in front of the participant with the center and top of monitor at level with the subject’s eyes. Each picture in an audiovisual trial was shown for 2,250 ms, which was longer than the audio file to allow timely judgment. Sounds were presented binaurally via inserted earphones (Etymotic Research ER-3A) at 60 dB sensation level calibrated according to an individual's hearing threshold (Rao et al., 2010). The inter-trial interval was 750 ms (Fig. 4.2).

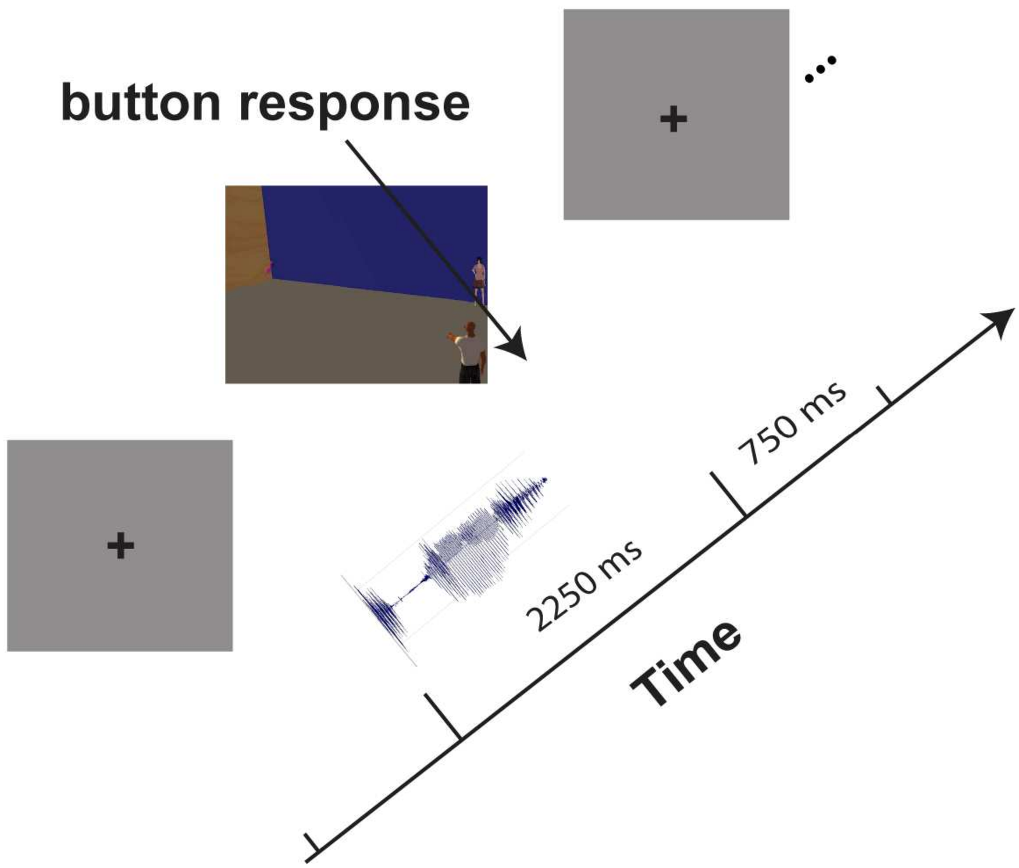

Fig. 4.2. Schema of simultaneous audiovisual presentation protocol.

The subjects sat comfortably in an acoustically and electrically treated room (Acoustic Systems) and spent approximately one hour to complete the experiment in four sessions with breaks. Instructions to the subject were that a male figure would be referring to an object (a flamingo, table, chair, or lamp) in the visual scene on the computer monitor. A male voice expressed either "this N" or "that N" (where N = flamingo, table, chair, or lamp) in the English paradigm or "kono N," "sono N," or "ano N" in the Japanese paradigm (where N = the Japanese translation of one of the four objects). Subjects were familiarized with the experimental task with a two-minute practice session, during which all the different types of audiovisual combinations were presented.

The varied parameters included three different relative spatial positions of speaker, hearer, and referent object; the presence or absence of a pointing gesture; and the demonstrative forms. Congruent pairings consisted of “this” or “kono” when the object was in speaker-near space, “that” or “sono” in hearer-near space, and “that” or “ano” for distant objects. Incongruent pairings consisted of all other arrangements. For example, if the male figure was pointing to a lamp beside him and the auditory stimulus was “this lamp” in English or “kono ranpu” in Japanese, the combination would be classified as a congruent pairing. If the auditory stimulus was instead “that lamp,” “sono ranpu,” or “ano ranpu,” the presentation would be classified as an incongruent pairing. The gesture parameter was not used to determine congruency but served as an independent variable.

The four trial conditions (congruent and incongruent in the presence of gesture, congruent and incongruent in the absence of gesture) were each presented 80 times in a random order. The subjects were instructed to press one of two keyboard buttons (DirectIN-PCB keyboard by Empirisoft Corp.) to indicate “yes” or “no” that the picture and the auditory verbal expression formed an expected pairing. Subjects were fitted with a 64-channel Waveguard cap, with shielded wires for its Ag/AgCl electrodes. The EEG cap contained a silicon ring for conductive gel for each electrode. The impedance of the electrodes was maintained at a maximum of 5 kOhm. The electrodes on the EEG cap were configured to match the International 10-20 Montage System and intermediate locations. The EEG data were recorded with the Advanced Neuro Technology system (http://www.ant-neuro.com/). The AFz electrode served as the ground. The common average of the connected unipolar electrodes of the ANT amplifier served as the default reference. The sampling rate was 512 Hz. The bandpass for EEG recording was set between 0.016 and 200 Hz. The ERP event markers were time-locked to the onset of the audiovisual presentation in each trial.

4.2.4. Behavioral data analysis.

Behavioral response and reaction times were recorded for each subject. Since the pairings of visual scene and demonstrative expression were classified as either a congruent or incongruent response according to the distance-based definitions in each language (Aoyama, 1995; Matsushita, 1901; Sakuma, 1936; Halliday & Hasan, 1979; Lyons, 1977; Quirk, 1979), participants' responses were counted in terms of conformity rate to the conventional usage. A "correct" response would therefore be agreement ("yes") with a congruent pairing or disagreement ("no") for an incongruent pairing. Mean reaction times for each condition were also calculated for each subject. Mixed repeated-measure ANOVA tests were performed to test three main factors and their interaction: subject group (English vs. Japanese), gesture (presence vs. absence), and congruency (congruent vs. incongruent). In the case of significant interactions, further two-tailed t-tests were applied for the factors of interest.

4.2.5. ERP data analysis.

As in the previous study (Stevens & Zhang, 2013), the off-line ERP analysis was performed with the ASA 4.7 package (Advanced Source Analysis, ANT Corp.) with common average reference. The choice of common average reference is in line with recommendation for EEG recordings with 64 or more electrodes (Curran, Tucker, Kutas, & Posner, 1993; Picton et al., 2000) and deemed appropriate for the analysis of N400 or P600 responses (Hamm, Johnson, & Kirk, 2002; Lau, Stroud, Plesch, & Philips, 2006; Lim, Padmala, & Pessoa, 2009), even though earlier studies using fewer electrodes typically employed the mastoid (or linked mastoids) reference. The ERP epochs were 900 ms long, including a 100 ms pre-stimulus baseline. Trials with peaks exceeded µ50 µV in amplitude were rejected as artifact. A bandpass filter (0.5~40 Hz) was applied. A minimum of 50 good trials per condition was required for a subject to be included in ERP averaging and statistical analysis.

To determine the time windows for the ERP components of interest, a Bonferroni-corrected point-to-point t-test was applied in comparing the congruent and incongruent conditions at the midline parietal Pz electrode (cf. Gunter, Friederici, & Schriefers, 2000; Kutas & Hillyard, 1980). An area analysis of selected time windows was chosen, rather than peak analysis, because the ERP responses of our current interest occurred in relatively late time windows and showed slow, broad deflections rather than sharp peaks (Stevens & Zhang, 2013).

The electrodes were grouped into nine electrode regions (Fig. 4.3). Similar regional groupings of electrodes have been used in previous ERP studies (Rao et al., 2010; Stevens & Zhang, 2013; Zhang et al., 2011). At the anterior scalp, a left frontal (LF) group included F4, F6, F8, FC6, C6, and TP8. A right frontal (RF) group included the electrodes opposite to the LF group. A frontal-mid-central (FMC) group included FC1, FC2, C3, and C4. A Left temporal-parietal (LTP) group TP8, CP6, CP4, P8, P6, and

89

P4, and a right temporal-parietal (RTP) group included the right hemisphere counterparts. A group of midline electrodes in the parietal region (MID) included CPz, CP1, CP2, Pz, P1, and P2. A left parietal-occipital (LPO) included PO8, PO6, PO4, and O2. A right parietal-occipital (RPO) group covered the opposite sites on the right hemisphere. A midline parietal-occipital (MPO) group included POz and Oz.

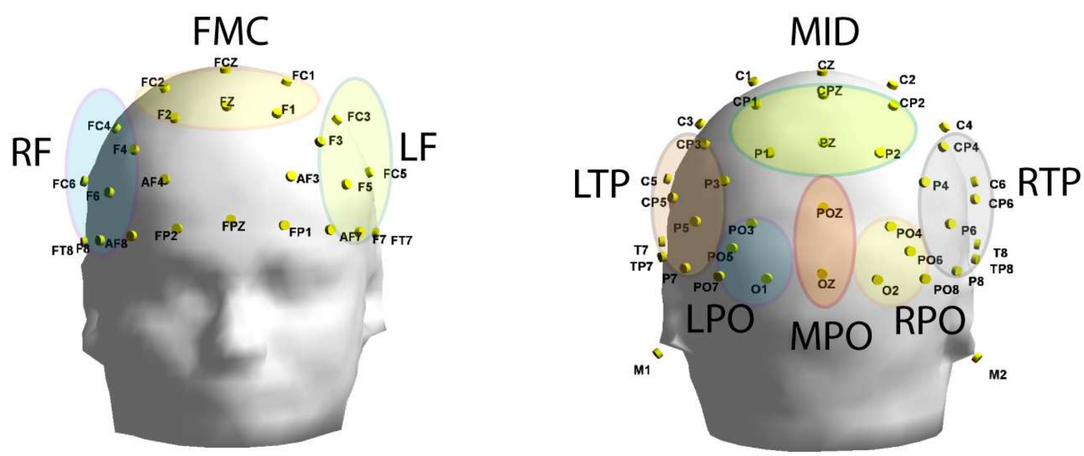

Fig. 4.3. Grouped electrode sites of interest. Sites are based on a realistic MRI-based head model.

A two-tailed point-to-point t-test with a Bonferroni correction was performed on the averaged ERP waveform data of the selected electrode sites. The t-test was applied to analyze differences between the congruent and incongruent conditions across time (Guthrie & Buchwald, 1991; Zhang et al., 2011). The responses were collapsed across the proximal and distal demonstratives respectively for the gesture condition and no-gesture condition so that the combined physical properties of the audio and visual stimuli were identical to avoid the confounding factor of physical stimulus difference in the congruent-incongruent comparison (Stevens & Zhang, 2013). An ERP congruency effect in this study would be defined as showing a significant difference ($p < 0.05$) that persisted for at least 40 ms (Zhang et al., 2011).

Additionally, ANOVA tests were performed using Systat (Version 12) statistics software with the following main factors: subject group, congruency, gesture, and electrode region. Post-hoc tests were performed for individual electrode regions when a congruency effect was found during a time interval. Laterality effects were tested by comparing right hemisphere sites (RF, RTP, RPO) and left hemisphere sites (LF, LTP, LPO). The Global Field Power (GFP) was calculated in order to consider overall differences in electric potential for all 64 electrodes at every sampling point of the epoch window (Hamburger & Burgt, 1991; Lehmann & Skrandies, 1980). Sampling points for both the congruent and incongruent conditions were translated into z scores relative to the 100 ms baseline GFP activity (Rao et al., 2010). Significant effects were taken into account when the differences ($p < 0.01$) persisted for at least 40 ms.

4.3. Results.

4.3.1. Behavioral data.

The reaction times showed consistent congruency effect in both subject groups regardless of the presence or absence of gesture. In the English data, incongruent presentations resulted in longer reaction times ($F(1,9) = 16.6$, $p < 0.01$). The Japanese reaction time data showed a similar congruency effect ($F(1,9) = 26.5$, $p < 0.001$). While there was no significant effect for the three main factors in the accuracy data, a significant three-way interaction of gesture x congruency x subject group was found ($F(1,18) = 21.7$, $p < 0.001$). The interaction of gesture and language was also significant ($F(1,18) = 11.3$, $p < 0.01$). Both English and Japanese participants showed consistent judgment (95% and above) in the congruent condition. Incongruent presentations with gesture also produced similar rejection rates in the two subject groups, approximately 70% for both groups. A language effect was found for the trials without an accompanying gesture, in which the English speakers showed a larger conformity rate difference between the congruent and incongruent trials than the Japanese speakers ($t = 2.37$, $p < 0.05$) (Fig. 4.4).

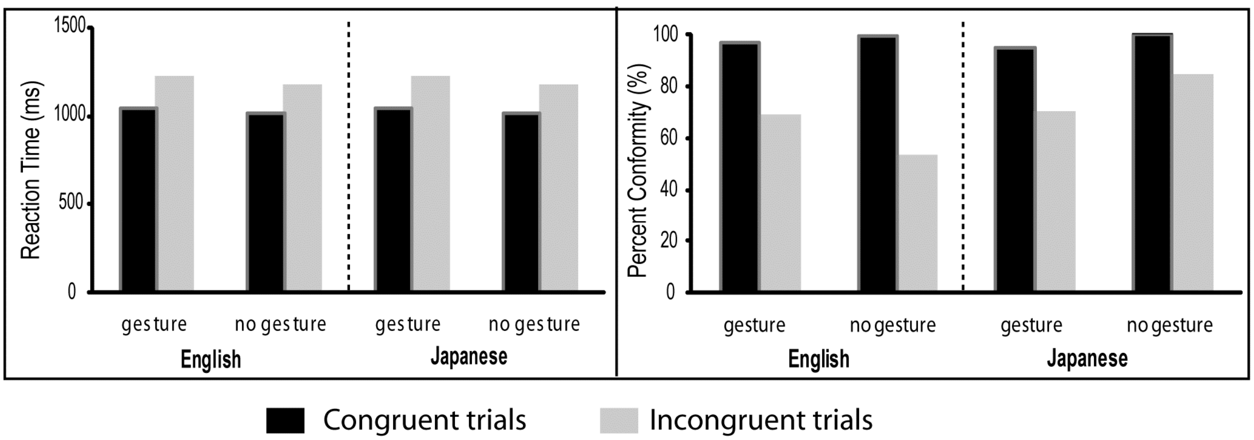

Fig. 4.4. Behavioral data showing reaction time and percentage of conformity. Data include congruent and incongruent trials in the two subject groups.

4.3.2. ERP data.

Three time windows were selected: 325-385 ms, 400-500 ms, and 500-775 ms, based on the significant differences in point-to-point t-tests performed on data from the Pz electrode. Mixed ANOVA tests included four main factors: language (or subject group), congruency, gesture, and electrode site. Planned comparisons in repeated-measures ANOVA were further performed for the two languages separately to identify contributions from each subject group.

In the 325-385 ms time interval, multiple significant interaction effects were found, including four-way interaction of subject group, congruency, gesture, and electrode site ($F(8,144) = 5.4$, $p < 0.01$). The three-way interaction of gesture x congruency x electrode site was also significant ($F(8,144) = 6.3$, $p < 0.001$), as was the interaction of subject group x gesture x congruency ($F(1,18) = 12.8$, $p < 0.01$) and the two-way interaction of gesture x congruency ($F(1,18) = 9.2$, $p < 0.01$).

The English and Japanese data were then tested separately with ANOVA. The English data showed a significant interaction of gesture and congruency ($F(1,9) = 16.2$, $p < 0.01$). Further analyses were performed for the gesture trials and *no-gesture* trials separately. In the gesture trials, the interaction between congruency and electrode site was significant when a gesture was present ($F(8,72) = 4.6$, $p < 0.001$). Post-hoc t-test analysis of the electrode sites showed a significant congruency effect at MPO with a negative deflection ($t(9) = -4.4$, $p < 0.01$) (Fig. 4.5). The Japanese data did not show significant effects in the early window of 325-385 ms for the gesture trials (Fig 4.6).

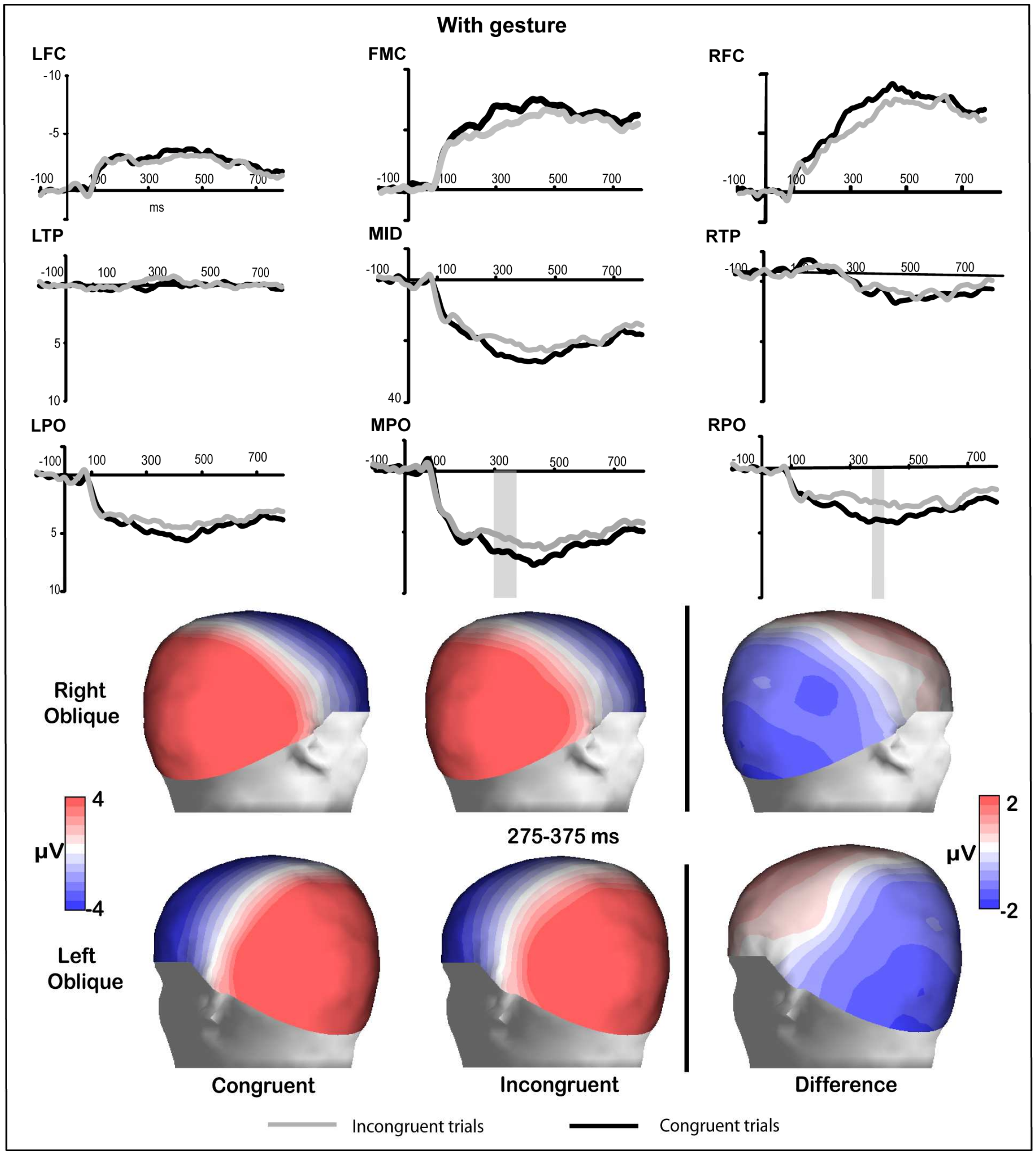

Fig. 4.5. Grand average English ERP data for gesture trials. Shaded areas indicate minimally 40 ms time intervals where the point-to-point t-test was significant [$p < 0.05$]. Topographical maps shown for congruent, incongruent, and difference between conditions.

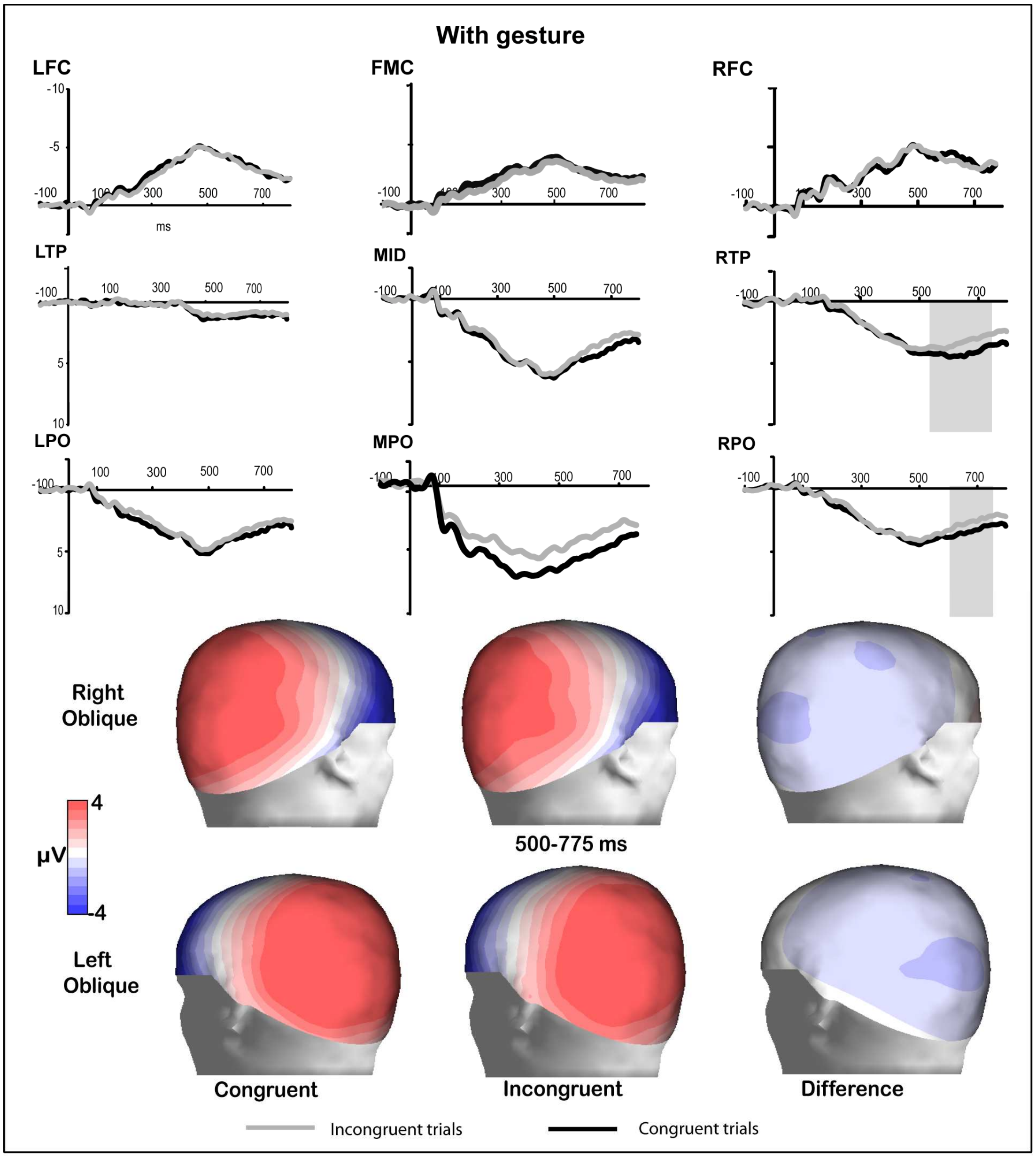

Fig. 4.6. Grand average Japanese ERP data for the gesture trials. Shaded areas indicate significant congruent vs. incongruent differences using the same convention as in Figure 4.5. Topographical maps shown for congruent, incongruent, and difference between conditions.

In the *no-gesture* trials in English, the congruency effect was significant, ($F(1,9) = 40.0$, $p < 0.001$). There was also an interaction of congruency x electrode site ($F(8,72) = 4.7$, $p < 0.001$). In particular, congruency effects were found at LFC ($t(9) = -4.8$, $p < 0.001$) and

RFC (t(9) = -3.5, p <.01). In these frontal channel groupings, the incongruent condition consistently has a more negative ERP response than the congruent condition. In the posterior channels, there were significant congruent effects at MID ($t(9) = -4.2$, $p < 0.01$) and MPO ($t(9) = -4.6$, $p < 0.01$) (Fig. 4.7). Like the gesture trials, the *no-gesture* trials in Japanese did not show any effect for demonstrative use for the 325-385 ms time window (Fig. 4.8).

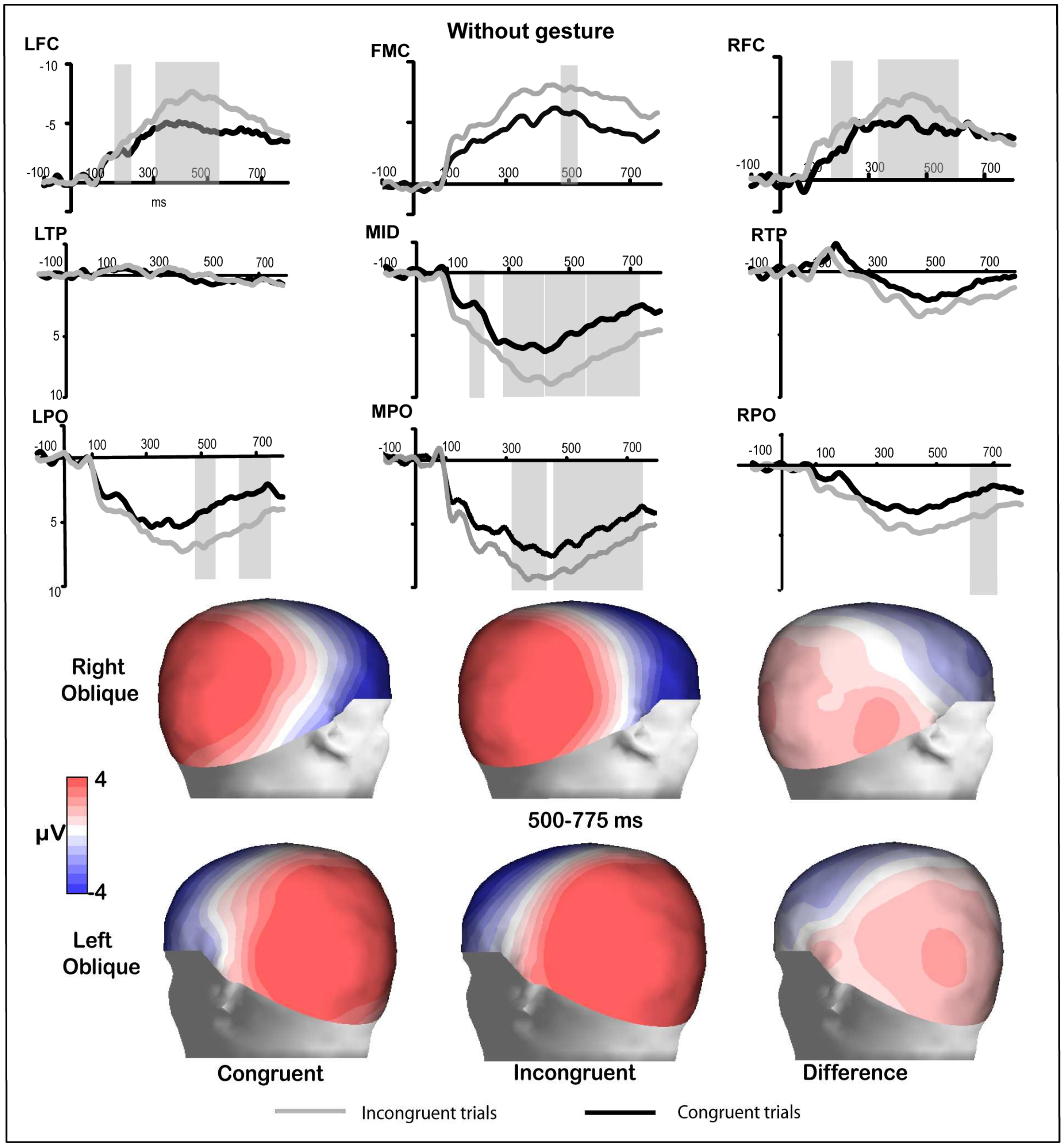

Fig. 4.7. Grand average English ERP for the no-gesture trials. Shaded areas indicate significant congruent vs. incongruent differences using the same convention as in Figure 4.5. Topographical maps shown for congruent, incongruent, and difference between conditions.

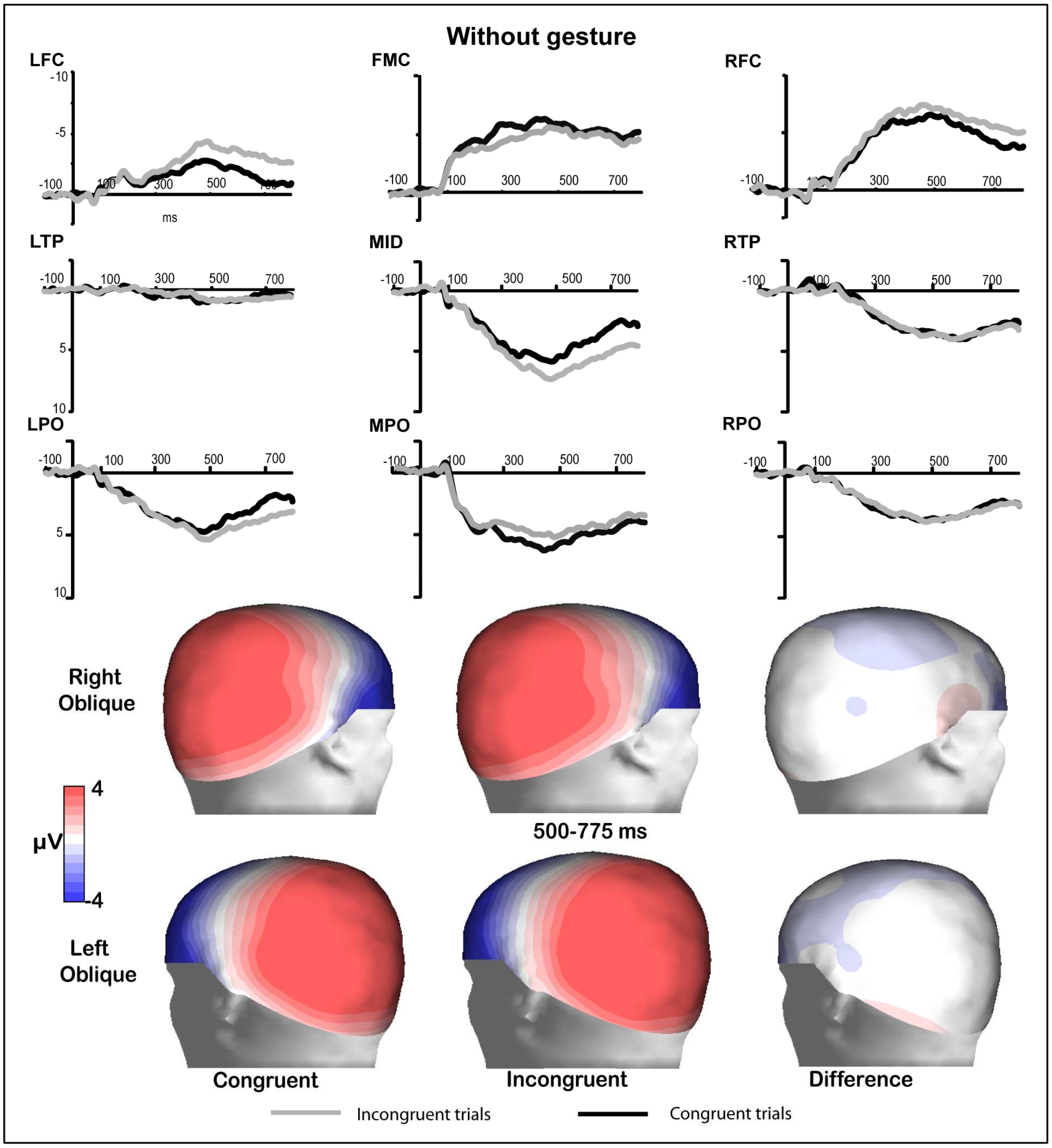

Fig. 4.8. Grand average Japanese ERP data for no-gesture trials. Topographical maps shown for congruent, incongruent, and difference between conditions.

In the 400-500 ms time window, the mixed ANOVA test showed significant interaction of gesture x congruency x electrode site ($F(8,144) = 4,5$, $p < 0.05$). The gesture trials showed no significant effects in either English or Japanese data. For the *no-gesture* trials, the English data alone showed a main congruency effect ($F(1,9) = 8.8$, $p < 0.05$). In particular, the congruency effect was found at LFC ($t(9) = -3.3$, $p < 0.01$), RFC

($t(9) = -2.7$, $p < 0.05$), FMC ($t(9) = -2.8$, $p < 0.05$), LPO ($t(9) = -2.7$, $p < 0.05$) with a more negative deflection for the incongruent pairings. A congruency effect with a positive deflection was found at MID ($t(9) = -2.5$, $p < 0.05$), and at MPO ($t(9) = -2.3$, $p < 0.05$) (Fig. 7). Like the gesture trials, the Japanese data for the no-gesture trials showed no significant effects in the 400-500 ms window (Fig. 4.8).

In the 500-775 ms time, mixed ANOVA test showed significant interactions of subject group x congruency x gesture x electrode site ($F(8,144) = 4.5$, $p < 0.05$) and subject group x congruency ($F(1,18) = 4.4$, $p < 0.05$). In the gesture trials, there was no significant effect in the English data. The Japanese data showed a congruency effect for the *gesture* trials ($F(1,9) = 4.9$, $p = 0.05$) (Fig. 4.6). The right hemisphere showed larger negative deflections than the left ($F(1,9) = 5.9$, $p < 0.05$). Site-specific tests showed the congruency effect at RTP ($t(9) = -4.2$, $p < 0.01$) and RPO ($t(9) = -2.9$, $p < 0.05$) for the trials that included a gesture for the Japanese subjects. In the *no-gesture* trials, only the English data showed a congruency effect with a positive deflection observed at MID ($t(9) = 3.0$, $p < 0.05$), LPO ($t(9) = -2.9$, $p < 0.05$), RPO ($t(9) = -3.4$, $p < 0.01$), and MPO ($t(9) = -3.6$, $p < 0.01$) (Fig. 4.7). The *no-gesture* trials did not show a congruency effect in the Japanese group (Fig. 4.8).

The GFP data showed time windows with significant differences between the congruent and incongruent conditions in the two subject groups (Fig. 4.9). In the gesture trials, significant differences were found among English-speaking participants at 262-322 ms and 398-473 ms. In contrast, Japanese speakers showed significant differences later at 500-551 ms, 596-645 ms, and from 717 ms. In the no gesture trials, the English data showed significant differences during the following intervals: 168-322 ms, 398-473 ms, 484-551 ms, and 586-645 ms, and the Japanese data showed a significant congruency effect during the 717-760 ms time window.

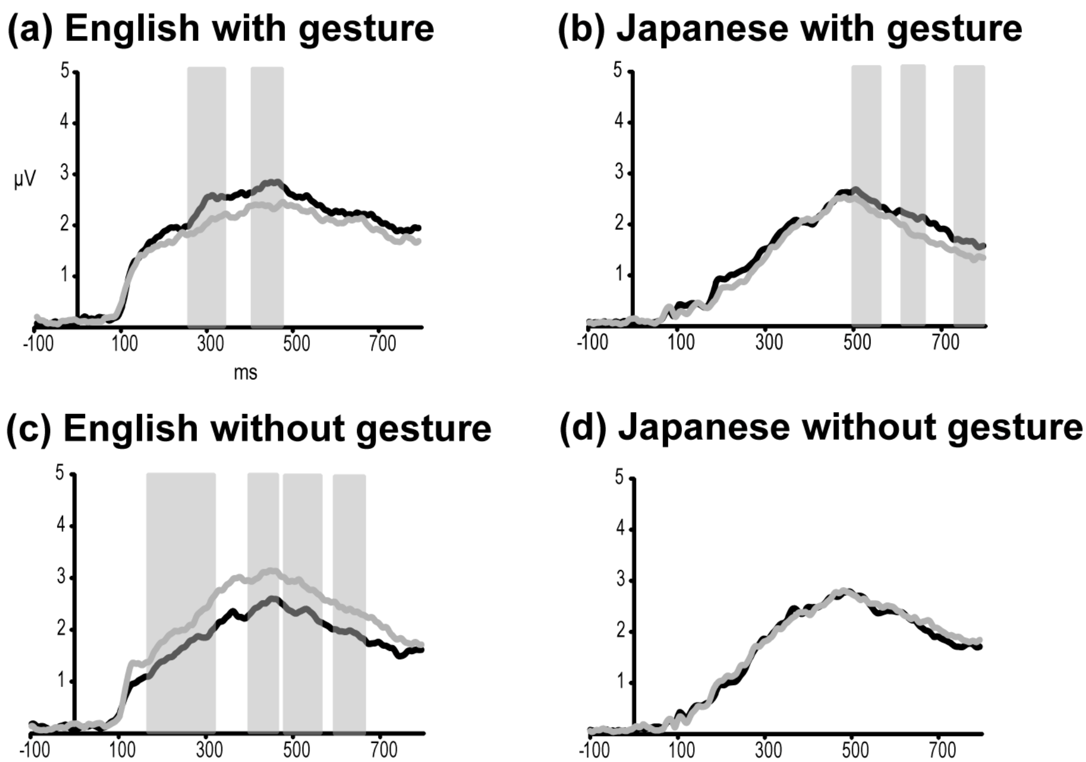

Fig. 4.9. Global Field Power data from the two subject groups and gesture conditions. Shaded areas indicate significant congruent vs. incongruent differences using the same convention as in Figure 4.5.

4.4. Discussion.

The perceptual experiment on the use of spatial demonstratives and gesture brings us closer toward understanding how humans interpret the physical environment when communicating thought. The present data offer new findings regarding the electrophysiological markers of embodied gestural processing and visual spatial processing in connection with referential expressions. Consistent with predictions and other published work (Stevens & Zhang, 2013), an N400-like effect was observed for the gesture trials and longer reaction times were recorded when demonstratives did not match participants' expectations for a given spatial configuration of discourse participants and referent object in both English and Japanese. When the pointing gesture was removed from the trials, English speakers' data showed a significant positive P600-like deflection for the incongruent trials. But no such significant ERP effect was observed the Japanese subject group although reaction times were increased in the incongruent trials for both subject groups regardless of the presence or absence of gesture. The cross-linguistic ERP data during the *no-gesture* trials suggest that speakers of separate languages may rely on extra-linguistic cues to different degrees, possibly because of the specificity expressed in the lexicon.

Of the different interpretations of gesture use, the data here are most consistent with the Mutually Adaptive Modalities Hypothesis and the Lexical Semantics Hypothesis. The data do not support the Free Imagery Hypothesis, which considers gestures to be "pre-linguistic" symbols. In support of the Mutually Adaptive Modalities hypothesis, it was found that speakers of different languages showed varied reliance on gesture. The data further suggest that language itself is one determinant in a user's expectation for a pointing gesture. Caution is necessary here as the theoretical accounts only make claims about how people produce gestures in certain contexts, not how they will perceive gestures in the same settings. If gesture comprehension (or expectation of gesture use) directly relies on one's knowledge of gesture production in the context of spatial demonstrative use, greater reliance on gestures in production would lead to increased expectation of gestural presence in the visual scenes in the present study.

Both subject groups consistently showed a significant congruency effect (i.e., differences between the congruent and incongruent conditions) in the behavioral data when a pointing gesture was used with the spatial demonstratives. The incongruent demonstrative use in both languages led to a longer reaction time, suggesting that the subjects expected the demonstrative use to conform to the proximal-distal distinction in English or the distance-based three-way *ko-so-a* distinction in Japanese. Additional support for the argument that English and Japanese speakers expect for demonstrative use to conform to some spatial configuration comes from their acceptability judgments. The congruent condition, that is, where the demonstrative form conformed to the binary proximal-distal distinction in English or the tertiary *ko-so-a* system in Japanese, was accepted by participants with a very high rate of “yes” judgment at over 95%. The rate of “no” judgment for the incongruent trials was more variable. Both subject groups rejected most incongruent trials when a gesture was present. In the no-gesture trials, Japanese speakers rejected incongruent audiovisual pairings to a high degree (84%), and English speakers rejected incongruent trials at a lower rate (53%). The lower rate of rejection among English speakers may suggest more difficulty or ambiguity in decision making. If English speakers rely on a co-occurring gesture in spatial reference resolution more than Japanese speakers, then the relatively lower rejection rate for English speakers could indicate that the English subjects did not have enough information to evaluate a demonstration without that act containing the overt pointing cue. The acceptability data for the no-gesture trials also indicate that English speakers showed a stronger bias toward classifying trials as congruent, something not observed among Japanese speakers’ conformity rates. Thus the proximal-distal distinction alone does not explain demonstrative use in English. The manner in which a pointing gesture affected participants’ decisions suggests gestures play an important role in demonstrative reference resolution. Japanese speakers, by contrast, may have been able to assess a similar linguistic act with sufficient information provided by their demonstrative lexicon and the spatial context without an additional indexical gesture. Together, these findings suggest that English and Japanese speakers rely on a pointing gesture differently to judge a referential act, such that this extra-linguistic tool appears more informative to English speakers to limit the possible domain of reference. The domains of usage of the demonstrative forms in English and Japanese did appear to correlate with the relative positions of the speaker and hearer, a finding consistent with model of control space, such that relative proximity would afford an agent more action possibilities toward an object. English and Japanese therefore have both lexical and extra-linguistic tools to delineate a referential space, and demonstrative forms coincided with discourse participants' position to interact with an object.

The ERP data supported the first prediction that unexpected pairings of demonstratives and visual scenes based on proximity in the presence of a pointing gesture would elicit an N400-like response for the congruency effect in both the English and Japanese subjects. As predicted according to the second hypothesis, there was also a language effect. The English data showed the N400-like effect after 300 ms in the midline posterior occipital region (MPO), and in the Japanese data the effect was in right posterior regions at a later latency. Although the Japanese data do not show a significant congruency effect at MPO, the scalp distributions are broadly similar in the two languages (Figs. 4.5 & 4.6). The GFP data confirmed the between-subject differences in the latency of the N400-like response for the congruency effect in the gesture trials (Figs. 4.5, 4.6, & 4.9). One plausible explanation would be that participants from the two language subject groups employed separate strategies to process the visual and auditory information online in developing an evaluation of the trial. For instance, Kelly, Kravitz, and Hopkins (2004) found an N400 waveform following an early sensory effect bilaterally in occipital and frontal sites when speech was presented with incongruent gestures. The authors conclude that hand gestures are integrated with spoken language at both early and late phases of processing. In the present study, speakers of Japanese and English may indeed interface differently with space to decide the appropriateness of a demonstrative expression with the pointing gesture, weighing elements of the spatial context differently due to the binary vs. tertiary lexical distinctions. The late broad negative waveform in the posterior electrode sites for the incongruent trials is consistent with previous N400 literature (Hagoort, Brown, & Swaab, 1996; Kutas & Hillyard, 1980; Lau, Stroud, Plesch, & Philips, 2008). Similar late N450 response has been reported for congruency effects involving iconic gestures (Wu & Coulson, 2007), and the study described in Chapter 3 also found a prolonged latency (Stevens & Zhang, 2013). Presumably, the task of integrating a spatial scene with semantic information may demand more cognitive resources than straight anomalies presented word by word in sentence form without any additional spatial processing. The ERP latency differences for the congruency effect, interestingly, did not directly translate into behavioral reaction time differences between the two subject groups. The online neural responses to incongruency may be a more sensitive indicator than the later motor action for decision making. Alternatively, the reaction time data and the ERP latency differences may not index the same processing stage or cognitive effort in time. The participants were required to make a “yes” or “no” decision if a demonstrative expression and a visual scene were matched up properly. The decision process involved responding to the naturalness of the physical stimuli, the linguistic act, and a categorical decision of trial acceptability. The richness of the stimuli likely led to variable ERP latencies at a different time window, which might not shift in tandem with behavioral reaction at a later window.

The ERP data revealed differences in the language factor (English vs. Japanese) as well as in the gesture factor (*no-gesture* vs. *gesture*). Unlike the posterior negative deflection for the *gesture* trials, the no-gesture ERP morphology and scalp distribution showed a different pattern for the congruency effect in the English data. The broad positive-going waveform at centroparietal sites that peaked approximately at post-stimulus 600 ms window indicated a P600-like effect (Hagoort et al., 1999; Kuperberg, 2007). The Japanese data, however, did not show such P600 congruency effect for the *no-gesture* trials despite the existence of the significant congruency effect at the behavioral level. The GFP data for the P600 effect were consistent with the waveform analysis (Fig. 4.9).

The P600 has been observed with ungrammatical syntactic construction and with semantic violations of verb-argument agreement as well as picture-sentence mismatch (Kuperberg, 2007; Vissers et al., 2008). In the present context, a positive deflection of the waveforms was observed in trials missing the pointing gesture for the English-speaking subject group. One interpretation is that demonstrative reference in English may depend on an expected structural pattern involving a linguistic form and physical context. The P600 has also been similarly observed for structural violations such as musical phrases that ended in out-of-key chords (Patel, Gibson, Ratner, Besson, & Holcomb, 1998). The structural conformity account could readily explain the generation of the P600 response, assuming that both speech and pointing gesture are integral components required by a multimodal embodied grammar.

Another plausible interpretation for the English-specific P600 in the present study is available with the Error Monitoring Model (van de Meerendonk, Kolk, Chwilla, & Vissers, C, 2009; van de Meerendonk, Kolk, Vissers, & Chwilla, 2010). According to this account, the P600 does not just reflect syntactic processing but instead is a general index of reanalysis. In this reanalysis process, the subject detects an error and attempts to edit the previous input to resolve a perceived conflict. The Error Monitoring Model for the "semantic P600" could offer an explanation for why the positive ERP effect was observed in a context (an incongruent demonstrative used to refer to an object without an accompanying gesture) that may require reanalysis in order to make a cognitive decision.

The language-specific P600 in relation to the absence of a pointing gesture during a demonstration suggests different processing mechanisms in the two populations of language users. The English subjects reanalyzed the visual input with an expected pointing gesture use to interpret the demonstration meaningfully. For the Japanese subjects, other contextual cues, such as joint gaze and relative distance in relation to the tertiary demonstrative system, may outweigh the pointing gesture in their congruency judgment under the present perceptual paradigm. The lack of a P600 response in the Japanese group suggests an influence from the language system on participants' perception and production of demonstrative language.

These cross-linguistic differences in the influence of gesture on speech comprehension indicate that speech and gesture influence each other to achieve the communicative goal. When that goal is demonstrative reference, the explicitness of information in one modality seems to compensate for under-specification by other means, a conclusion in line with the Mutually Adaptive Modalities hypothesis (De Ruiter, 2006). Pertinently for the use of demonstratives, the variable socio-linguistic reliance on a pointing gesture between English and Japanese speakers suggests that possibilities for interaction with an object can be understood by language users both by lexical as well as other physical means.

Lastly, of unknown significance, a negative ERP waveform difference between the congruent and incongruent conditions was observed in the English data before 350 ms. This early bilateral effect was accompanied with a late P600-like effect specific to the English subjects. The frontal electrode sites showed a negative deflection, and the central electrodes showed polarity reversal. The scalp distribution of the early effect did not resemble topographical patterns of vertical or horizontal eye movement. It is also different from a P2 effect in terms of waveform morphology and scalp topography, which was previously reported in tasks of image and language processing (Federmeier & Kutas, 2002). The latency of the early ERP response was similar to that of the ELAN (Early Left Anterior Negativity) (Friederici & Frisch, 2000). However, the ELAN is known more to track grammatical violations than semantic or contextual anomalies and is lateralized to the left. We can speculate that the early congruency effect in the present study could reflect differential sensory processing of the complex visual-spatial computation of the relative positions of objects for the congruent and incongruent trials in the absence of the pointing gesture.

The limitations of the study are important to appreciate, including the small sample size of speakers of English and Japanese. A larger sample size of subjects would be helpful to increase the power of the study. In future studies, it would be helpful to test speakers of other languages with different demonstrative systems. Coventry et al. (2008) have found, for example, that Spanish and English speakers use the proximal demonstrative ("este," "this") differently in each language. Also, English and Chinese show many differences in the use of spatial demonstratives, even though both languages use the binary demonstrative system (Wu, 2004). To further understand the cross-linguistic commonalities and differences in the use of deictic expressions and nonverbal gestures in communication, it would be helpful to study with similar methods languages with unique demonstrative systems, such as Malagasy, which includes expressions for different spatial contexts depending on whether the object is visible.

Another limitation of the present project has to do with the difficulty of reconstructing naturalistic referential act for social interaction in an experimental setting. There is lack of ecological validity in the experimental design (Brewer, 2000; Locke, 1986). The conclusions are drawn regarding linguistic intuitions from participants engaged in unnatural activities: making binary acceptability decisions by striking a computer key, looking at static images of gesturing instead of the normal dynamic motion, observing a demonstration divorced from greater discourse, and especially seeing the artificial setup of combining a demonstrative with a visual scene that does not have any accompanying gesture. Future work should aim for more naturalistic stimuli and settings including the animated gestures. Practically, aspects of the experimental design in a laboratory setting must differ from the typical real life experiences as strict experimental control needs to be exercised at the expense of realism. ERP studies have historically made use of abstract and de-contextualized experiment materials in a paradigm of expectancy violation. For example, a P600 experiment for syntactic processing may present words one at a time on a screen and string together strange word sequences that one almost never encounters in real life. Nevertheless, the violation paradigm remains the dominant, time-tested method in ERP research and has revealed important findings about the brain mechanism underlying cognitive, linguistic and social processes.

Basic questions regarding the nature of demonstrative reference can be pursued in an experimental setting, such as measuring the electrophysiological responses to unexpected demonstrative use according to the setting or an accompanying gesture. A definitive answer is not supplied by the present initial work in English and Japanese, but the results here highlight some of the complexity underlying an everyday and essential part of verbal communication. In the last chapter, which follows, I will briefly discuss the nature of a demonstrative as a single-use name, synthesize the findings presented thus far, and suggest new paths for such work and implications in clinical settings.

Chapter 5: Conclusion

Naming for control and disposal

5.1. On the disposable nature of demonstratives.

If we are to entertain the concept that the use of a demonstrative implies a speaker's possible actions toward an object, then a subsequent question is what unique properties would be true of demonstratives, the forms of which suggest the ranking of control spaces by which one could interact with an object. I suggest that demonstratives possess a ***disposable*** property, a quality not shared (or shared much less) by other referring expressions, which allows for the release of control over an object in order to refer to new objects with the same word. The *disposable* property of demonstratives means that the use of a demonstrative is a singular event, after which the use of the same form would not be expected to refer to the same object. Therefore, the 'disposal' of the demonstrative's reference involves a new sense of focus and specification with each use of the same expression. The control spaces associated with demonstratives are highly dependent on context and are dynamic sources of meaning.

The disposable property of demonstratives is derived from being a deictic use of language. Like other deictic language, demonstratives refer to a dynamic world. This distinguishes demonstratives from eternal names, e.g., Beatrice, which suggest a static world. Eternal names can be used to refer multiple times to an object imagined to be the same across contexts. If we were to refer to a baby rabbit as "Beatrice," the name will continue to refer to the animal the next day, as an adolescent, and after a fulfilled life as an elderly rabbit: we pretend that the object is somehow identical across instances of time. If we use the term “this” to refer to a baby rabbit, use of the term “this” the next day would not in all likelihood refer to the same object. Demonstratives can be thought of therefore as disposable names, as each use of a demonstrative introduces an object into a new symbolic relationship. In other words, each demonstration is a new naming event.

Demonstratives and other names, however, are used similarly in counterfactual situations. Counterfactual situations require that an object persist across possible worlds, e.g., “If <u>that</u> [*pointing to cup*] had been glass, it would have shattered right into the toys.” We easily entertain contradictory thoughts where an object simultaneously possesses one property (a cup being completely made of plastic) and another, mutually exclusive property (an imagined cup made of glass) for demonstratives and other named objects. A requirement of object persistence presents a problem for the logical treatment of natural language, especially in the use of possible worlds as a model for modal language. A referring expression that picks out an object that is understood to persist in all possible worlds is a *name* or, as Kripke also called it, a *rigid designator* (Kripke, 1972). Although the objects are *expected* to persist, our cup example shows the logical impossibility of allowing a named object to inhabit a different possible world and yet be considered the same object (that is, for a plastic cup to be glass).

Because demonstratives refresh their denotation with each use, we can use the same form repeatedly to refer to different objects in a way that other referring expressions do not produce equally felicitous utterances. To illustrate, let’s imagine a child escorted into a candy store by a grandparent eagerly indifferent to the child’s parents’ wishes, where we might hear (24), whereas (25) and (26) are less expected, as the child prances around the store pointing to different candies (with subscripts to identify different referents).

(24) I want this$_i$ and this$_j$ and this$_k$ and this$_l$ and this$_m$.

(25) ??I want it$_i$ and it$_j$ and it$_k$ and it$_l$ and it$_m$.

(26) ??I want the candy$_i$ and the candy$_j$ and the candy$_k$ and the candy$_l$ and the candy$_m$.

A nearly opposite pattern can be seen when a speaker speaks continuously about a single referent. Imagine that we are visiting a friend and walking the perimeter of a newly bought house. Consider how relatively felicitously the friend's utterances (27), (28), and (29) serve to describe the new purchase.

(27) ??We got this house for a decent price, and this house has good space, and this house is near amenities.

(28) We got the house for a decent price, and the house has good space, and the house is near amenities.

(29) We got it for a decent price, and it has good space, and it is near amenities.

(27) appears difficult to process, as the use of the demonstrative phrase seems to suggest that the referent is changing. (28) seems to cause less difficulty in understanding that the referent of "the house" is constant. Of the three utterances, (29) sounds to this author as the most natural and (27) the least. Notably, even if the speaker were gesturing to the house, this would not significantly improve the acceptability of (27). The examples (24)-(29) are offered to illustrate how demonstratives are used as disposable names, such that

the same referent is not expected to serve as the denotation for a demonstrative after a singular use. As such, they should follow a general naming rule. Kripke (1972) described the practice of naming as a sort of baptism.

The ritualistic nature of naming (and thus demonstrative use) will be described here according to three rules: an *index rule* (30), an *assignment rule* (31), and a *naming rule* (32). By positing these three rules, a naming event will be understood to require three actions: the use of an object (e.g., an arrow) to indicate some set of properties of interest; assignment of a name to that set of properties; and acknowledgement that the isolated properties now exist as a unique object, respectively.

The index rule serves the function of associating the referent with another object. It starts with an attempt to isolate some set of properties, Q, by forming a unique relationship between it and any object that can serve as an index. It therefore posits some function, F, which can have any meaning whatsoever, as long as it connects the indexical object uniquely with the set of properties of interest and excludes any other set of properties.

(30) The index rule.

Let *i* be some entity; and let *Q* and *R* be sets of properties; and let *Index* be a function of the form *Index*(*i*,*Q*,*F*), such that there exists some function *F* for which the following is true:

$$\exists i,Q,F[[F(Q) \rightarrow i] \wedge [F(i) \rightarrow Q] \wedge \forall R[Q \neq R \rightarrow \neg[[F(R) \rightarrow i] \wedge [F(i) \rightarrow R]]]$$

The function that relates a set of properties *Q* and the indexical object *i* can be imagined like a string that connects them, and one can start on either end of the string and follow it to the other end. The index rule serves to describe the necessity of having some instrument (physical, cognitive, etc.) to isolate the set of properties as a figure out of a nameless ground.

Another rule that is necessary for a naming event is the assignment of the name itself. We can consider a rule, (31), where some isolated set of properties, now a referent, will be identified by a name that will serve as a constant.

(31) Assignment rule

Let Q be some set of properties, "name" a property; let *Assign* be a function such that for any Q, *Assign*(Q,"name") will bind sign "name" to set Q such that for any function, F("name")$\rightarrow Q$

After having defined a rule to isolate a referent and assign this referent with a name, an additional step involves combining the *Index* rule with *Assign* rule to posit the existence of a uniquely identified object, one that has the exclusive property of having been named at any iteration of the *Naming* rule, (32).

(32) Naming rule

Let Q and R be sets of properties, "name" a property, and i an entity, such that for any set Q and entity i, if the function *Index* returns as true some unique relationship between Q and i and if the function *Assign* identifies a property (a name) with the set Q, then the existence of an object x and property q will be posited and *q* is only true for *Q*.

$$\forall Q,R,i[\text{index}\ (i,Q,R) \wedge \text{Assign}(Q,\text{“name”}) \rightarrow \exists x,q[q(Q) \wedge q(x) \wedge \neg q(R)]]$$

The effect of the naming rule in (32) is that when some set of properties is assigned a name, one can consider that an object has been formed and that that object has the unique property of being named by the rule. The existence of an object, therefore, will depend on the role of an observer to separate it from the rest of the universe by giving it a name.

The reason for positing a general naming rule such as (32) is that this should serve to form the basis of a taxonomy of more specific rules for the demonstrative forms in a language. An integral part of the naming rule is that some object serve as an index with which to isolate the referent. The nature of this object was not specified in (32), as it is assumed that different languages will make use of a diverse number of indices.

Demonstratives are argued here to be temporary names, and the indexical objects for demonstratives would include the associated control space. We could therefore notate one index as the first control space, the one highest ranked by the speaker, as *i*=1’. The index as the second control space can be notated as *i*=2’ and is the next highest ranked by the speaker and perhaps the highest estimated for the hearer, if the hearer does not share the speaker’s vantage point. If the referent is so distant in space or remote in memory that neither the point of view of the speaker nor that of the nearest available point of view seems adequate, the referent might be indexed by way of a third, non-local control space (*i*=3’). English demonstratives are therefore catalogued according to the rules in (33), and the Japanese forms in (34).

(33)

Rule for ‘this’

a. $\forall$Q,R[index (1’,Q,R) $\wedge$ Assign(Q,“this”) $\rightarrow$ $\exists$x,q[q(Q) $\wedge$ q(x) $\wedge$ $\neg$q(R)]]

Rules for ‘that’

b. $\forall$Q,R[index (2’,Q,R) $\wedge$ Assign(Q,“that”) $\rightarrow$ $\exists$x,q[q(Q) $\wedge$ q(x) $\wedge$ $\neg$q(R)]]

c. $\forall$Q,R[index (3’,Q,R) $\wedge$ Assign(Q,“that”) $\rightarrow$ $\exists$x,q[q(Q) $\wedge$ q(x) $\wedge$ $\neg$q(R)]]

Rule for ‘ko’

d. $\forall$Q,R[index (1’,Q,R) $\wedge$ Assign(Q,“ko”) $\rightarrow$ $\exists$x,q[q(Q) $\wedge$ q(x) $\wedge$ $\neg$q(R)]]

Rule for ‘so’

e. $\forall$Q,R[index (2’,Q,R) $\wedge$ Assign(Q,“so”) $\rightarrow$ $\exists$x,q[q(Q) $\wedge$ q(x) $\wedge$ $\neg$q(R)]]

Rule for ‘a’

f. $\forall$Q,R[index (3’,Q,R) $\wedge$ Assign(Q,“a”) $\rightarrow$ $\exists$x,q[q(Q) $\wedge$ q(x) $\wedge$ $\neg$q(R)]]

It should be stressed when reading the rules in (33) that there appears to be a level of symmetry between English and Japanese forms that should not be exaggerated. The forms in both languages are indeed related to ranked control spaces in the above naming rules, but the boundaries of the control spaces are not necessarily expected to be universal. In other words, the boundaries of the second control space may not be the same in English and Japanese and indeed should not be expected to be so, considering that Japanese includes a unique form for the second control space, whereas English shares a form for the second and third control spaces. The problem of dividing the control spaces is therefore reminiscent of classifying colors in English and Russian. The rules above defining demonstratives incorporate a level of cultural differences in the delineation of a control space. We should expect for differences between language communities, including reliance on gesture and gaze, to be considered implicitly by the above rules. Language communities will differ on their use of extra-linguistic communication, and we would predict that the lexical divisions will be inversely proportional to the richness of the non-lexical means to establish a control space. The taxonomy above invites investigation into other possible divisions, available in other languages. It would be worthwhile to consider new experiments that would address the possibility of more elaborate demonstrative systems and tabulate any corresponding poverty of physical constraints.

5.2. Synopsis.

This dissertation has sought to investigate normative behavior associated with demonstrative use. Demonstrative forms have been described as conforming to the relative distance to the speaker (Anderson & Keenan, 1985; Bain, 1879; Greenberg, 1985; Lyons, 1977), but they have also been argued to relate to the attentional state of the hearer (Fitzgerald, 1966; Kirsner & van Heuven, 1988; Kirsner, 1977; Strauss, 1993, 2002). The two studies described in this dissertation project used an ERP technique that showed clear evidence that while both claims are true, there are aspects of the demonstrative use that are not fully accountable by either claim.

Due to the perceived shortcomings in previous descriptive models of demonstratives, a new approach has been proposed that demonstrative forms relate to how the speaker ranks possible actions toward the referent object. The new proposal invokes the idea of control space and argues that a speaker will demonstrate an object with an implied suggestion for how the object could be controlled. A space is defined, with such tools as a pointing gesture and gaze, in which certain actions may be calculated, including grasping, approaching, etc. The concept of a ranking of control spaces motivating demonstrative form selection resembles a description of demonstratives in terms of a *control cycle* (Brovold & Grush, 2012; Langacker, 2002). The proposal here differs from the control cycle, which does not address languages with more than two demonstrative forms and treats control as a categorical property. The current proposal instead includes an array of ranked control spaces, with correlated forms, in order to accommodate languages, such as Japanese, with more than two demonstrative forms.

The concept of control spaces led to three hypotheses that were considered here. To evaluate the hypotheses, ERP studies were conducted that simultaneously presented visual scenes of a virtual world environment that included a male figure, a female figure, and a referent object with an auditory stimulus that included a demonstrative expression. Participants were asked to decide if the presented audio-visual scenes formed congruent pairings of visual context with an expected demonstrative form or incongruent pairings. Participants' judgments and reaction times were collected, and the EEG signal was averaged over each condition to yield the event-related potentials (ERPs). In light of the previous literature on semantic anomaly and surprising contexts, we expected that incongruent visual-audio pairings would elicit an N400 component, a negative waveform deflection at about 400 ms after the stimulus.

Since languages have different divisions of control spaces and associated forms, we would hypothesize that language users would depend on extra-linguistic information, such as in form as a pointing gesture, according to such properties of their individual language as the number of demonstrative forms. This claim extrapolates from the Mutually Adaptive Modalities Hypothesis, which views language and gesture to be used in compensatory concert for the sake of effective communication (De Ruiter, 2006; Melinger & Levelt, 2005). The hypothesis of the individual language affecting reliance on gesture use was borne out by the ERP study. In the setting of an incongruent pairing of a demonstrative expression and a visual context that does not include an accompanying pointing gesture, the English subject group showed a late positive deflection in the ERP analysis that is understood to represent a P600 form, an ERP component observed in several settings that show a structural violation in the stimulus pattern (Hagoort et al., 1993, 1999; Osterhout & Holcomb, 1992; Schlesewsky & Bornkessel, 2006; Van Herten et al., 2006). The finding of cross-linguistic variation in response to a demonstration without gesture may suggest that English speakers rely more considerably on extra-linguistic cues to interpret a reference. The ERP results support a view that individual languages function differently in combination with other tools of communication.

A second hypothesis, with a nod toward the traditional distance-based model of demonstrative use, claims that relative distance of the referent object to the speaker and hearer will affect form selection, because different action possibilities will be estimated for discourse participants on the basis of how far they are from the object. Behavioral data strongly confirmed the view that distance affects demonstrative form selection in both English and Japanese. Participants conformed to this expectation in their acceptability judgments, and reaction times support the view that the congruency of a demonstrative form and a visual context depended on the relative distances of the referent to the speaker and hearer. The ERP data corroborated the hypothesis, as well, as visual contexts that included a pointing gesture provoked a late negative deflection in the waveform when the verbal stimulus did not coincide with the expected distances between interlocutors and referent. The negative waveform deflection is interpreted here to be akin to the N400 component observed in studies that presented stimuli with odd contexts (Osterhout, Allen, Mclaughlin, & Inoue, 2002; Brown & Hagoort, 1993; Kutas & Federmeier, 2011; Kutas & Hillyard, 1980, 1984).

Another hypothesis qualifies the second, as relative distance is expected to be less important in form selection if the hearer does not share the speaker's gaze. Giving credence to the role of attention in demonstrative selection, the ERP data did not reveal an N400 effect when the hearer's gaze was turned elsewhere but did show an N400 if the speaker and hearer were attending in the same direction. The importance of gaze has clear ramifications in terms of the function of demonstrative forms. Attentional models (e.g, Fitzgerald, 1966; Kirsner & van Heuven, 1988; Strauss, 1993, 2002) stressed the role of certain forms, e.g. "this," to signal high focus. According to Strauss (2002), the so-called proximal form in fact signals a referent of high importance to the speaker and new information to the hearer. The data from this project support a role for attention, for which the hearer's gaze served as a proxy. When the hearer is not attending in the same direction as the speaker, the speaker would likely assess fewer action possibilities between the hearer and the referent at that moment, even if the referent is physically nearer the hearer.

What unifies both the hypotheses and results from the experiments discussed above is an approach to word selection in referring to an object that implies how the speaker plans to interact with the object. I suggest that the proposed approach offers a more grounded account for language use, such that conceptual divisions between proximal and distal demonstrative forms are derived from the interaction of the language users' bodies with their environment. The proposal here seeks to explain demonstration in language as an example of embodied cognition, such that interpretation of the world depends on the subject's own condition. To the extent that an individual has tendencies to interact with the environment, an assessment of control of that environment ensues. Demonstratives are argued here to identify one's potential actions toward an object by means of contrastive lexical items that co-occur with extra-linguistic cues. The incorporation of non-verbal communication establishes a relationship between mature language use and non-human animal communication, as well as child language acquisition. We have suggested and provided evidence for the claim that the richness of a language's demonstrative system affects language users' dependence on physical cues, a conclusion consistent with the Mutually Adaptive Modalities Hypothesis (De Ruiter, 2006; Melinger & Levelt, 2005).

There are several ways to extend the argument. As mentioned earlier, there are languages with a similar number of demonstrative forms as English (e.g., Chinese) or Japanese (e.g., Spanish) that would be expected to show differences in the use of such forms. There are languages such as Malagasy that include many more demonstrative forms than English. There are languages, such as French, where the deictic element of demonstratives can be dropped. That is, to refer to this cup or that cup in French, one can say "cette tasse-cì" or "cette tasse-là," respectively, yet the final morpheme can be omitted. To what extent this would affect French speakers' reliance on extra-linguistic cues differently than the previously discussed languages would be informative.

Another extension of control spaces would be in the temporal use of demonstratives. The demonstrative "this" is used to refer either the present or the near future, whereas “that” can be used for other views of time, such as toward the past, e.g., (34).

(34) a. Back when I met your dad, he was always talking about deep sea fishing during the summer. Does he actually do that?

b. <u>That</u> summer, yes, to a fault, but definitely not <u>this</u> summer.

b’. <u>This</u> summer, yes, to a fault, but definitely not <u>that</u> summer.

The use of “this” in (34) is understood to refer to a present summer or one soon to arrive, no matter the order of presentation, that is, (34b) or (34b’). This interpretation can be explained as an extension of control space to the description of time. No matter a speaker’s power of nostalgia, one has more possible actions available toward the future than the past.

A fuller understanding of demonstratives would benefit from developmental research. Young children have been shown to not easily take other individuals’ perspectives (de Villiers & de Villiers, 1974; Rodrigo et al., 2004), yet non-egocentrism does not seem to be a requirement of mature use of “this” and “that” according to experimental data (Webb & Abrahamson, 1976). Children might not show any differences in ERP data in the interpretation of demonstrative use. Alternatively, the congruency effect in children might in fact be stronger than in adults, as children may apply more rigid rules in evaluating demonstratives than adults (Landry & Loveland, 1989). As mentioned in Chapter 2, children developmentally learn to demonstrate an object by releasing direct control over an object. Further work on this subject would therefore expect children to over-generalize one demonstrative form in a language and develop the mature distribution in tandem with their delayed recognition of social cues. Children would similarly be expected to develop language-specific spatial terms in conjunction with their physical capacity to interact with their environment more fully.

Clinical applications for the present project may include therapies for individuals who may differ from normative behavior in terms of their capacity to understand others' perspectives. Two populations that may show a deficit in reading the intended actions in others include those on the autism spectrum (Hobson, García-Pérez, & Lee, 2010; Kanner, 1943; Landry & Loveland, 1989) and those with schizophrenia (Brüne, 2003; Corcoran, Mercer, & Frith, 1995). Individuals with autism and schizophrenia are at special risk to misinterpret communicative gestures. Joint attention serves as a prognostic indicator of language acquisition in the autistic population (Bruinsma, Koegel, & Koegel, 2004). Gaze tracking and joint attention have been associated with higher-level domain-general meta-representation and executive function (Stone & Gerrans, 2006). Therapeutic techniques that incorporate a speaker's use of control spaces in referential language may benefit individuals who may be challenged with reference resolution, abstract representation, and indeed interpreting others' body language.

A larger question that this topic approaches is the capacity for different individuals to analyze the social dynamic of a communicative context. The control space framework is built upon the premise that cognition develops in relation to one's interaction with the environment. A specific prediction for populations that may differ in their capacity for social interpretation, e.g., autistic individuals, is that these persons would be expected to develop improved mastery over demonstrative distribution in proportion to their own physical skills. It has already been reported that many physical skills are developed through imitation. Thus, those individuals who are better able to interact physically with their surroundings would be likely to benefit more from imitation. By learning the affordances of their environment, these individuals would more readily acquire mature demonstrative use, if not broader principles of linguistic expression as applied to a dynamic context. If the above prediction were borne out by experiment, then a therapeutic intervention for communicative skills in autistic individuals would be to encourage physical skills development through imitation.

References

Anderson, S., & Keenan, E. (1985). Deixis. In T. Shopen (Ed.), *Language Typology and syntactic description, Volume III* (p. 259–398). Cambridge: Cambridge University Press.

Aoyama, T. (1995). Deixis and value: A semantic analysis of the Japanese demonstratives. In E. Contini-Morava, B. Sussman Goldberg, & R. Kirsner (Eds.), *Meaning as explanation: advances in linguistic sign theory* (pp. 289–319). Berlin: Walter de Gruyter.

Arbib, M. (2002). The mirror system, imitation, and the evolution of language. In C. Nehaniv & K. Dautenhahn (Eds.), *Imitation in animals and artifacts* (pp. 229–280). Cambridge, MA: MIT Press.

Armstrong, D., & Wilcox, S. (2007). *The gestural origin of language*. New York: Oxford University Press.

Austin, J. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.

Bain, A. (1879). *A higher English grammar*. Henry Holt and Company.

Baron-Cohen, S. (1996). *Mindblindness: an essay on autism and theory of mind*. Boston: MIT Press/Bradford Books.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.

Bates, E., Camaioni, L., & Volterra, V. (1975). Performatives prior to speech. *Merrill-Palmer Quarterly, 21*, 205–226.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science, 8*, 492–499.

Bjoertomt, O., Cowey, A., & Walsh, V. (2002). Spatial neglect in near and far space investigated by repetitive transcranial magnetic stimulation. *Brain*, *125*, 2012–2022.

Blaschke, M., & Ettlinger, G. (1987). Pointing as an act of social communication in monkeys. *Animal Behaviour*, *35*, 1520–1523.

Bloom, L. (1970). *Language development: Form and function in emerging grammars*. Cambridge, MA: MIT Press.

Boddaert, N., Chabane, N., Belin, P., Bourgeois, M., Royer, V., Barthelemy, C., Zilbovicius, M. (2004). Perception of complex sounds in autism: abnormal auditory cortical processing in children. *American Journal of Psychiatry*, *161*, 2117–2120.

Bortel, R., & Sovka, P. (2007). Regularization techniques in realistic Laplacian computation. *IEEE Transactions on Biomedical Engineering*, *54*, 1993–1999.

Bowerman, M. (1978). The acquisition of word meaning: An investigation into some current conflicts. In M. Foster and S. Brandes (Eds.), *Symbol as sense: New approaches to the analysis of meaning* (pp. 277–299). New York: Academic Press.

Bowerman, M. (1996). Learning how to structure space for language: a cross-linguistic perspective. In P. Bloom (Ed.), *Language and space* (pp. 385–436). Cambridge, MA: MIT Press.

Bowerman, M. (2007). Containment, support, and beyond: Constructing topological spatial categories in first language acquisition. In M. Aurnague, M. Hickmann, & L. Vieu (Eds.), *The Categorization of Spatial Entities in Language and Cognition* (pp. 177–203). Amsterdam: John Benjamins.

Brewer, M. (2000). Research design and issues of validity. In H. Reis & C. Judd (Eds.), *Handbook of research methods in social and personality psychology*. Cambridge: Cambridge University Press.

Brooks, R. (1990). Elephants don't play chess. In P. Maes (Ed.), *Designing Autonomous Agents*. Cambridge, MA: MIT Press.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Brovold, A., & Grush, R. (2012). Towards an (improved) interdisciplinary investigation of demonstrative reference. In Raftopoulos and Machamaer, (Eds.), *Perception, realism, and the problem of reference* (pp. 11-42). Cambridge, UK: Cambridge University Press.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44.

Brüne, M. (2003). Social cognition and behaviour in schizophrenia. In M. Brüne, H. Ribbert, & W. Schiefenhövel (Eds.), *The Social Brain. Evolution and Pathology* (pp. 277–313). Chichester, UK: Wiley and Sons.

Bruinsma, Y., Koegel, R., & Koegel, L. (2004). Joint attention and children with autism: a review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*, *10*(3), 169–175.

Butterworth, B., & Hadar, U. (1989). Gesture, speech and computational stages: A reply to McNeill. *Psychological Review, 96*, 168–174.

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 9–33). Mahwah, NJ: Lawrence Erlbaum Associates.

Byron, D., & Stoia, L. (2005). An analysis of proximity markers in collaborative dialogs. In *Proceedings of the Chicago Linguistic Society*, *41*(2), 17–32.

Call, J., & Tomasello, M. (1994). Production and comprehension of referential pointing by orangutans (Pongo pygmaeus). *Journal of Comparative Psychology*, *108*, 307–317.

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The role of declarative pointing in developing a theory of mind. *Infancy, 5*, 291–308.

Caminiti, R., Ferraina, S., & Johnson, P. (1996). The sources of visual information to the primate frontal lobe: a novel role for the superior parietal lobule. *Cerebral Cortex*, *6*, 319–328.

Caprici, O., Iverson, J., Pizzuto, E., & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child Language*, *23*, 645–673.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4).

Chou, T., Chen, C., Wu, M., & Booth, J. (2009). The role of inferior frontal gyrus and inferior parietal lobule in semantic processing of Chinese characters. *Experimental Brain Research*, *198*, 465–475.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. MIT Press.

Clark, E. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65–110). New York: Academic Press.

Corcoran, R., Mercer, G., & Frith, C. (1995). Schizophrenia, symptomatology and social inference: Investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research*, *17*(1), 5–13.

Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., & Núñez, R. (2009). Gesture and metaphor comprehension: Electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and Cognition*, *70,* 42–52.

Coventry, K., Valdes, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near-far perceptual space and spatial demonstratives. *Cognition*, *108*, 889–895.

Cowey, A., Small, M., & Ellis, S. (1994). Left visuo-spatial neglect can be worse in far than in near space. *Neuropsychologia*, *32,* 1059–1066.

Cowey, A., Small, M., & Ellis, S. (1999). No abrupt change in visual hemineglect from near to far space. *Neuropsychologia*, *37*, 1–6.

Curran, T., Tucker, D., Kutas, M., & Posner, M. (1993). Topography of the N400: Brain electrical activity reflecting semantic expectancy. *Electroencephalography and Clinical Neurophysiology*, *88*(3), 188–209.

De Gelder, B., Böcker, K., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from face and voice: Early interaction revealed by human electric brain responses. *Neuroscience Letters*, *260*, 133–136.

De Ruiter, J. (2006). Can gesticulation help aphasic people speak, or rather, communicate? *Advances in Speech-Language Pathology*, *8*(2), 124–127.

De Villiers, P., & de Villiers, J. G. (1974). On this, that, and the other: non-egocentrism in very young children. *Journal of Experimental Child Psychology*, *18*, 438–447.

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *Neuroscientist*, *13*, 580–593.

Delgado, B., Gómez, J., & Sarriá, E. (1999). Non-communicative pointing in preverbal children. In *Paper presented at the IXth European Conference on Developmental Psychology, Spetses, Greece.*

Demuynck, K., Duchateau, J., Van Compernolle, D., & Wambacq, P. (2000). An efficient search space representation for large vocaculary continuous speech recognition. *Speech Communication*, *30*(1), 37–53.

Desrochers, S., Morissette, P., & Ricard, M. (1995). Two perspectives on pointing in infancy. In In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in devleopment* (pp. 85–101). Hillsdale, NJ: Erlbaum.

Dick, A., Goldin-Meadow, S., Hasson, U., Skipper, J., & Small, S. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping*, *30*(11), 3509–3526.

Diessel, H. (1999). *Demonstratives: Form, function, and grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, *17*, 463–489.

Enfield, N. (2001). "Lip-pointing": A discussion of form and function with reference to data from Laos. *Gesture*, *1*(2), 185–211.

Enfield, N. (2003). Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language*, 82–117.

Epstein, R., Harris, S., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, *23*, 115–125.

Epstein, R., & Kanwisher, N. (1998). The cortical representation of the local visual environmnet. *Nature*, *392*, 598–601.

Farnè, A., & Làdavas, E. (2002). Auditory peripersonal space in humans. *Journal of Cognitive Neuroscience*, *14*(7), 1030–43.

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia*, *40*, 730–747.

Filimon, F., Nelson, J., Hagler, D., & Serenoa, M. (2007). Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *Neuroimage*, *37*, 1315–1328.

Fillmore, C. (1971). Santa Cruz Lectures on Deixis. In *Mimeo*. Indiana University Linguistics Club.

Fitzgerald, J. (1966). *Peirce's Theory of Signs as a Foundation for Pragmatism*. The Hague: Mouton.

Friederici, A., & Frisch, S. (2000). Verb argument structure processing: the role of verb-specific and argument-specific information. *Journal of Memory and Language, 43*(3), 476–507.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.

Gamberini, L., Seralgia, B., & Priftis, K. (2008). Processing of peripersonal and extrapersonal space using tools: Evidence from visual line bisection in real and virtual environments. *Neuropsychologia, 46*(5), 1298–1304.

Gibson, J. (1977). The concept of affordances. In *Perceiving, acting, and knowing* (pp. 67–82).

Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Goldin-Meadow, S., & Butcher, C. (2003). Pointing toward two-word speech in young children. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 85–107). Mahwah, NJ: Lawrence Erlbaum Associates.

Gredebäck, G., Melinder, A., & Daum, M. (2010). The development and neural basis of pointing comprehension. *Social Neuroscience*, *5*(5-6), 441–450.

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., & Kircher, T. (2009). Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Human Brain Mapping*, *30*, 3309–3324.

Greenberg, J. (1985). Some iconic relationships among place, time, and discourse deixis. In *Iconicity in syntax* (pp. 271–287).

Grindrod, C., Bilenko, N., Myers, E., & Blumstein. (2008). The role of the left inferior frontal gyrus in implicit semantic and selection: an event-related fMRI study. *Brain Research*, *1229*, 167–178.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274–307.

Gundel, J.K., & Johnson, K. (2013). Children's use of referring expressions in spontanous discourse: implications for theory of mind development. *Journal of Pragmatics*, *56*, 43-57.

Gundel, J.K., Ntelitheos, D., & Kowalsky, M. (2007). Children's use of referring expressions: some implications for Theory of Mind. *ZAS Papers in Linguistics*, *48*, 1-21.

Gunter, T., Friederici, A., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, *12*(4), 556–568.

Guthrie, D., & Buchwald, J.(1991). Significance testing of difference potentials. *Psychophysiology*, *28*, 240–244.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. In S. Garnsey (Ed.). In *Language and Cognitive Processes. Special Issue: Event-Related Brain Potentials in the Study of Language*, *Volume 4* (pp. 439–483). Hove: Lawrence Erlbaum Associates.

Hagoort, P., Brown, C., & Osterhout, L. (1999). *The neurocognition of syntactic processing*. New York: Oxford University Press.

Hagoort, P., Brown, C., & Swaab, T. Y. (1996). Lexical-semantic event-related potential effects in patients with left hemisphere lesions without lesions. *Brain, 119,* 627–649.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438–441.

Halliday, M., & Hasan, R. (1979). *Cohesion in English*. London: Longman Group Limited.

Halligan, P., & Marshall, J. (1991). Left neglect for near but not far space in man. *Nature*, *350*, 498–500.

Hamburger, H., & Burgt, M. (1991). Global field power measurement versus classical method in the determination of the latency of evoked potential components. *Brain Topography*, *3*, 391–396.

Hamm, J., Johnson, B., & Kirk, I. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, *113*(8), 1339–1350.

Hanks, W. (1992). The indexical ground of deictic reference. In A. Duranti & C. Goodwin (Eds.), *Rethinking Context: Language as an Interactive Phenomenon* (pp. 46–75). Cambridge, UK: Cambridge University Press.

Heilman, K., Watson, R., Valenstein, E., & Damasio, A. (1983). Localization of lesions in neglect. In A. Kertesz (Ed.), *Localization in Neuropsychology* (pp. 471–492). New York: Academic Press.

Hess, J., Novak, M., & Povinelli, D. (1993). "Natural pointing" in a rhesus monkey, but no evidence of empathy. *Animal Behaviour*, *46*, 1023–1025.

Hickmann, M. (2007). Static and dynamic location in French: Developmental and cross-linguistic perspectives. In M. Aurnague, M. Hickmann, & L. Vieu (Eds.), *The Categorization of Spatial Entities in Language and Cognition* (pp. 205–231). Amsterdam: John Benjamins.

Himmelmann, N. (1992). Demonstratives in narrative discourse: a taxonomy of universal uses. In *Studies in anaphora* (pp. 205–254). Amsterdam: John Benjamins.

Hobaiter, C., & Byrne, R. (2012). Gesture in consortship: wild chimpanzees' use of gesture for an "evolutionarily urgent" purpose. In S. Pika & K. Liebal (Eds.), *Developments in Primate Gesture Research* (pp. 129–146). Amsterdam/Philadelphia: John Benjamins.

Hobson, R., García-Pérez, R., & Lee, A. (2010). Person-centred (deictic) expressions and autism. *Journal of Autism and Developmental Disorders*, *40*(6), 653–664.

Holle, H., & Gunter, T. (2007). The role of iconic gesture in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*, 1175–1192.

Hollich, G., Hirsch-Pasek, K., Golinkoff, R., Brand, R., Brown, E., Chung, H., & Boroditsky, L. (2000). Breaking the language barrier: an emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, i–135.

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *NeuroReport*, *7*, 2325–2330.

Iverson, J., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*, 368–371.

Johnson, B., & Hamm, J. (2000). High-density mapping in an N400 paradigm: evidence for bilateral temporal lobe generators. *Clinical Neurophysiology*, *111*, 532–545.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Kelly, S., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, *89*, 252–260.

Kelly, S., Manning, S., & Rodak, S. (2008). Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass*, *2*, 569–588.

Kelly, S., Ward, S., Creigh, P., & Bartoletti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language*, *101*, 222–233.

Kemmerer, D. (1999). “Near” and “far” in language and perception. *Cognition*, *73*, 35–63.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–225.

King, B. (2004). *The dynamic dance: nonverbal communication in African great apes*. Cambridge, MA: Oxford University Press.

Kirsner, R. (1977). Deixis in discourse: an exploratory quantitative study of the Modern Dutch demonstrative adjectives. In T. Givón (Ed.), *Discourse and syntax* (pp. 355–375). New York: Academic Press.

Kirsner, R., & van Heuven, V. (1988). The significance of demonstrative position in Modern Dutch. *Lingua*, *76*, 209–248.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture Window into thought and action* (pp. 162–185). Cambridge, UK: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*, 16–32.

Krauss, R., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hands gestures tell us? In M. Zanna (Ed.), *Advances in experimental social psychology, Volume 28* (pp. 389–450). Tampa: Academic Press.

Krauss, R., Chen, Y., & Gottesman, R. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261-283). Cambridge: Cambridge University Press.

Krauss, R., Dushay, R., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, *31*, 533–552.

Kripke, S. (1972). Naming and Necessity. In D. Davidson & G. Harman (Eds.), *Semantics of Natural Language* (pp. 763–769). Dordrecht: Reidel.

Küntay, A., & Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? *Journal of Child Language*, *33*(2), 302–320.

Kumashiro, M., Ishibashi, H., Itakura, S., & Iriki, A. (2002). Bidirectional communication between a Japanese monkey and a human through eye gaze and pointing. *Current Psychology of Cognition*, *21*(1), 3–32.

Kuperberg, G. (2007). Neural mechanisms of language comprehension: Challenges to syntax, *Brain Research* (1146), 23–49.

Kuperberg, G., Sitnikova, T., Caplan, D., & Holcomb, P. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*, 117–129.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potential reflect semantic incongruity. *Science, 207*, 203–205.

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, R. (1974). Remarks on “this” and “that.” In *Proceedings of the Chicago Linguistic Society, Volume 10* (pp. 345–456). Chicago.

Lancaster, J., Woldorff, M., Parsons, L., Liotti, M., Freitas, C., Rainey, L., & et al. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, *10*(120-131).

Landau, B., & Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*, 217–238.

Landry, S. H., & Loveland, K. A. (1989). The effect of Social-Context on the Functional Communications-Skills of Autistic Children. *Journal of Autism and Developmental Disorders*, *19*, 283–299.

Langacker, R. (1987). *Foundations of cognitive grammar: Theoretical prerequisites, Volume 1.* Stanford University Press.

Langacker, R. (2002). The control cycle: why grammar is a matter of life and death. In *Proceedings of the Second Annual Meeting of the Japanese Cognitive Linguistics Association* (pp. 193–220).

Langacker, R. (2009). *Investigations in cognitive grammar*. Berlin: Walter de Gruyter.

Lau, E., Stroud, C., Plesch, S., & Philips, C. (2006). The role of prediction and lexical frequency effects in sentence processing. *Brain and Language*, *98*, 74–88.

Lau, E., Stroud, C., Plesch, S., & Philips, C. (2008). A cortical network for semantics: (de)constructing the N400. *Brain and Language*, *98,* 74–88.

Leavens, D., & Hopkins, W. (1999). The whole hand point: The structure and function of pointing from a comparative perspective. *Journal of Comparative Psychology*, *113*, 417–425.

Leavens, D., Hopkins, W., & Bard, K. (1996). Indexical and referential pointing in chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *110*, 346–353.

Leavens, D., Hopkins, W., & Thomas, R. (2004). Referential communication by chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *118*(1), 48–57.

Leavens, D., Russell, J., & Hopkins, W. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (Pan troglodytes). *Child Development*, *76*, 291–306.

Legrand, D., Brozzoli, C., Rossetti, Y., & Farnè, A. (2007). Close to me: multisensory space representation. *Consciousness and Cognition*, *16*, 687–699.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology*, *48,* 609–621.

Leonard, R. (1995). Deixis in Swahili: attention meanings and pragmatic function. In E. Contini-Morava and B. Goldberg (Eds.), *Meaning as explanation: advances in sign-based linguistics* (pp. 271–287). Berlin: Mouton de Gruyter.

Levinson, S. (1996). Frames of reference and Molyneux's question: Cross-linguistic evidence. In P. Bloom (Ed.), *Language and space* (pp. 109–169). Cambridge, MA: MIT Press.

Levinson, S. (2003). *Space in Language and Cognition*. Cambridge: Cambridge University Press.

Lim, S., Padmala, S., & Pessoa, L. (2009). Segregating the significant from the mundane on a moment-to-moment basis via direct and indirect amygdala contributions. *Proceedings of the National Academy of Science (USA)*, *106*, 16841–16846.

Liszkowski, U. (2006). Infant pointing at twelve months: Communicative goals, motives, and social-cognitive abilities. In N. Enfield & S. Levinson (Eds.), *Roots of Human Sociality: Culture, cognition, and interaction* (pp. 153–178). New York: Berg.

Locke, E. (1986). Generalizing from laboratory to field: Ecological validity or abstraction of essential elements. In *Generalizing from laboratory to field studies* (pp. 3–9).

Lucy, J. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lyons, J. (1977). *Semantics, Volume 2*. New York: Cambridge University Press.

Macmillan, N., & Creelman, C. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Majid, A., Bowerman, M., Kita, S., Haun, D., & Levinson, S. (2004). Frames of reference and language concepts. *Trends in Cognitive Science*, *8*, 108–114.

Masataka, N. (2003). From index-finger extension to index-finger pointing: Ontogenesis of pointing in preverbal infants. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 69–84). Mahwah, NJ: Lawrence Erlbaum Associates.

Matsushita, D. (1901). *A dictionary of colloquial Japanese*. Tokyo: Benseisha.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D., & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 141–161). Cambridge, UK: Cambridge University Press.

Melinger, A., & Levelt, W. (2005). Gesture and communicative intention of the speaker. *Gesture*, *4*(2), 119–141.

Moore, C., & D'Entremont, B. (2001). Developmental changes in pointing as a function of attentional focus. *Journal of Cognition and Development*, *2*, 109–129.

Moss, H., Rodd, J., Stamatakis, E., Bright, P., & Tyler, L. (2005). Anteromedial temporal cortex supports fine-grained differentiation among objects. *Cerebral Cortex*, *15*, 616–627.

Naito, E., Scheperjans, F., Eickhoff, S., Amunts, K., Roland, P., Zilles, K., & et al. (2008). Human superior parietal lobule is involved in somatic perception of bimanual interaction with an external object. *Journal of Neurophysiology*, *99*, 695–703.

Nieuwland, M., & Van Berkum, J. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Brain Research (Cognitive Brain Research), 24*, 691–701.

Nöe, A., & Thompson, E. (2004). Are there neural correlates of consciousness? *Journal of Consciousness Studies, 11*(1), 2–28.

Norman, D. (1988). *The psychology of everyday things*. New York: Basic Books.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Osterhout, L., & Holcomb, P. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.

Osterhout, L., Holcomb, P., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experiment Psychology: Learning, Memory, & Cognition*, *20*, 786–803.

Osterhout, L., Willems, R., Kita, S., & Hagoort, P. (2002). Brain potentials elicited by prose-embedded linguistic anomalies. *Memory & Cognition*, *30*, 1304–1312.

Özyürek, A., Kita, S., Allen, S., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures? Insights from crosslinguisitic variations and similarities. *Gesture*, *5*(1/2), 219–240.

Özyürek, A., Willems, R., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*, 605–616.

Paczynski, M., Kreher, D., Ditman, T., & Holcomb, P. (2006). Electrophysiological evidence for the role of animacy and lexico-semantic association in processing nouns within passive structures, *Cogntive Neuroscience Supplement, Abstract*.

Pascual-Marqui, R., Esslen, M., Kochi, K., & Lehmann, D. (2002). Functional imaging with low resolution brain electromagnetic tomography (LORETA): review, new comparisons, and new validation. *Japanese Journal of Clinical Neurophysiology*, *30*, 81–94.

Pascual-Marqui, R., Michel, C., & Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology*, *18*, 49–65.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience*, *10*(6), 717–733.

Perlman, M., Tanner, J., & King, B. (2012). A mother gorilla's variable use of touch to guide her infant: insights into iconicity and the relationship between gesture and action. In S. Pika and K. Liebal (Eds.), *Developments in Primate Gesture Research* (pp. 55–71). Philadelphia, PA: John Benjamins.

Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, S.,Taylor, M. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, *37*, 127–152.

Pika, S., Liebal, K., Call, J., & Tomasello, M. (2005). The gestural communication of apes. *2005*, *5*(1/2), 41–56.

Pika, S., Liebal, K., & Tomasello, M. (2003). Gestural communication in young gorillas (Gorilla gorilla): Gestural repertoire, learning, and use. *American Journal of Primatology*, *60*, 95–111.

Pizzuto, E. (2002). Communicative gestures and linguistic signs in the first two years of life. In *Paper presented at the EURESCO conferences, Brain Development and Cognition in Human Infants*. Maratea, Italy, June 7-12, 2002.

Pizzuto, E., & Capobianco, M. (2005). The link and differences between deixis and symbols in children's early gestural-vocal system. *Gesture*, *5*(1/2), 179–199.

Ponelis, F. (1993). *Development of Afrikaans*. Frankfurt: Peter Lang.

Povinelli, D., & O'Neil, D. (2000). Do chimpanzees use their gestures to instruct other? In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: perspectives from developmental cognitive neuroscience* (pp. 459–487). New York: Oxford University Press.

Povinelli, D., & Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends in Cognitive Science*, *7*, 157–160.

Proverbio, A., & Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*, *459*, 142–146.

Quirk, R. (1979). *A grammar of Contemporary English*. London: Longman Group Limited.

Racine, T. (2012). Cognitivism, adaptationism, and pointing. In S. Pika and K. Liebal (Eds.), *Developments in Primate Gesture Research, Volume 6* (pp. 165–180). Philadelphia, PA: John Benjamins.

Rao, A., Zhang, Y., & Miller, S. (2010). Selective listening of concurrent auditory stimuli: An event-related potential study. *Hearing Research*, *268*, 123–132.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. A Bradford Book.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Language within Our Grasp*, *21*(5), 188–194.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology*, *12*, 149–154.

Rodrigo, M. J., Gonzalez, A., de Vega, M., Muneton-Ayala, M., & Rodriguez, G. (2004). From gestural to verbal deixis: a longitudinal study with Spanish infants and toddlers. *First Language*, *24*, 71–90.

Sadehipour, A., & Kopp, S. (2011). Embodied gesture processing: motor-based perception-action integration in social artificial agents. *Cognitive Computation*, *3*, 419–435.

Sakuma, K. (1936). *Expressions and usages in modern Japanese*. Tokyo.

Savage-Rumbaugh, E. S., McDonald, K., Sevcik, R., Hopkins, W., & Rupert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (Pan paniscus). *Journal of Experimental Psychology: General*, *115*, 211–235.

Schegloff, E. (1984). On some gestures' relation to talk. In J. Atkinson and J. Heritage (Eds.), *Structures of social action* (pp. 266–296). Cambridge: Cambridge University Press.

Schlesewsky, M., & Bornkessel, I. (2006). Context-sensitive neural responses to conflict resolution: Electrophysiological evidence from subject-object ambiguities in language comprehension. *Brain Research*, *1098*, 139–152.

Schmitt, B., Münte, T., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology*, *37*, 474–484.

Sekihara, K., Sahani, M., & Nagarajan, S. (2005). Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction. *Neuroimage*, *25*, 1056–1067.

Semino, E., & Culpeper, J. (2002). *Cognitive stylistics: language and cognition in text analysis*. Amsterdam: John Benjamins.

Simos, P., Breier, J., Maggio, W., Gormley, W., Zouridakis, G., Willmore, L., & et al. (1999). Atypical temporal lobe language representation: MEG and intraoperative stimulation mapping correlation. *Neuroreport*, *10*, 139–142.

Sitzman, S. (2013, July 3). Gesturing as we speak. *New York Times*. Retrieved from http://www.nytimes.com/2013/07/04/opinion/gesturing-as-we-speak.html

Sommerville, J., & Decety, J. (2006). Weaving the fabric of social interaction: articulating developmental psychology and cognitive neuroscience in the domain of motor cognition. *Psychonomic Bulletin & Review*, *13*, 179–200.

Sperber, D., & Wilson, D. (2005). Pragmatics. In F. Jackson & M. Smith (Eds.), *Oxford Handbook of Contemporary Philosophy* (pp. 468–501). New York: Oxford University Press.

Stevens, J., & Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: An event-related potential study. *Journal of Neurolinguistics*, 26(1): 31-45.

Stevens, J. & Zhang, Y. (2014). Brain mechanisms for processing co-speech gesture: A cross-language study of spatial demonstrattives. *Journal of Neurolinguistics*, 30, 27-47.

Stone, V. E., & Gerrans, P. (2006). What's domain-specific about theory of mind? *Social Neuroscience*, *1*, 309–319.

Strauss, S. (1993). Why "this" and "that" are not complete without "it." In K. Beals, G. Cooke, D. Kathman, K.E. McCullough, S. Kita, and D. Testen (Eds), *CLS 29: Papers from the 29th Regional Meeting of the Chicago Linguistics Society* (pp. 403–17). Chicago.

Strauss, S. (2002). This, that, and it in spoken American English: a demonstrative system of gradient focus. *Language Sciences*, *24*, 131–152.

Sutton, S., Braren, M., Zubin, J., & John, E. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *155*, 1436–1439.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical Publishers.

Tanner, J., & Byrne, R. (1996). Representation of action through iconic gesture in a captive lowland gorilla. *Current Anthropology*, *37*, 162–173.

Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Thrane, T. (1980). *Referential-semantic analysis. Analysis of a theory of linguistic reference*. Cambridge: Cambridge University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2006). Why don't apes point? In N. Enfield & S. Levinson (Eds.), *Roots of Human Sociality: Culture, cognition, and interaction* (pp. 506–524). Oxford & New York: Berg.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M., Call, J., Nagell, K., Olguin, K., & Carpenter, M. (1994). The learning and use of gestural signals by young chimpanzees: a trans-generational study. *Primates*, *35*, 137–154.

Tomasello, M., Call, J., Warren, J., Frost, G., Carpenter, M., & Nagell, K. (1997). The ontogeny of chimpanzee gestural signals: a comparison across groups and generations. *Evolution of Communication*, *1*, 223–259.

Tomasello, M., George, A., Kruger, A., Farrar, J., & Evans, A. (1985). The development of gestural communication in young chimpanzees. *Journal of Human Evolution*, *14*, 175–186.

Tychonoff, A., & Arsenin, V. (1977). *Solution of ill-posed problems*. Washington: Winston & Sons.

Ullsperger, P., Erdmann, U., Freude, G., & Dehoff, W. (2006). When sound and pictures do not fit: Mismatch negativity and sensory integration. *International Journal of Psychophysiology*, *59*, 3–7.

Van Berkum, J., Zwitserlood, P., Hagoort, P., & Brown, C. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*, 701–718.

Van de Meerendonk, N., Kolk, H., Chwilla, D., & Vissers, C. (2009). Monitoring in language perception. *Language and Linguistics Compass*, *3*, 1211–1224.

Van de Meerendonk, N., Kolk, H., Vissers, C., & Chwilla, D. (2010). Monitoring language perception: mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, *22*, 67–82.

Van Herten, M., Chwilla, D., & Kolk, H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, *18*(7), 1181–1197.

Véa, J., & Sabater-Pi, J. (1998). Spontaneous pointing behaviour in the wild pygmy chimpanzees (Pan paniscus). *Folia Primatologica*, *69*, 289–290.

Vissers, C., Kolk, H., van de Meerendonk, N., & Chwilla, D. (2008). Monitoring in language perception: evidence from ERPs in a picture-sentence matching task. *Neuropsychologia*, *46*(4), 967–982.

Volterra, V., Caselli, M., & Capirci, O. (2005). Gesture and the emergence and development of language. In M. Tomasello & D. Slobin, (Eds.), *Beyond nature-nuture - Essays in honor of Elizabeth Bates* (pp. 3–40). Mahwah, NJ: Lawrence Erlbaum Associates.

Volterra, V., & Erting, C. (1990). *From gesture to language in hearing and deaf children*. Berlin/New York: Springer Verlag.

Vygotsy, L. (1978). *Mind in society: The development of higher psychosocial processes.* In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, (Eds.). Cambridge, MA: Harvard University Press.

Webb, P., & Abrahamson, A. (1976). Stages of egocentrism in children's use of "this" and "that": A different point of view. *Journal of Child Language*, *3*, 349–367.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*, 655–684.

Werner, H., & Kaplan, B. (1963). *Symbol formation*. New York: Wiley.

West, W., & Holcomb, P. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, *13*, 363–375.

Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In S. Kita (Ed.), *Pointing: where language, culture, and cognition meet* (pp. 171–215). Mahwah, NJ: Lawrence Erlbaum Associates.

Winawer, J., Witthoft, M., Frank, M., Wu, L., Wade, A., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination (pp. 7780–7785). Presented at the National Academy of Sciences.

Woodruff, G., & Premack, D. (1979). Intentional communication in the chimpanzee - development of deception. *Cognition*, *7*(4), 333–362.

Wu, Y. (2004). *Spatial demonstratives in English and Chinese*. Amsterdam: John Benjamins.

Wu, Y., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, *42*, 654–667.

Wu, Y., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain and Language*, *101*, 234–245.

Xitco, Jr., M., Gory, J., & Kuczaj, S. (2001). Spontaneous pointing by bottlenose dolphins (Tursiops truncatus). *Animal Cognition*, *4*, 115–123.

Xu, J., Gannon, P., Emmorey, K., Smith, J., & Braun, A. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Science (USA)*, *106*(49), 20664–9.

Zhang, Y., Kuhl, P., Imada, T., Iverson, P., Pruitt, J., Stevens, E., & et al. (2009). Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. *Neuroimage*, *46*, 226–240.

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., Carney, E. (2011). Neural coding of formant-exaggerated speech in the infant brain. *Developmental Science*, *14*(3), 566–581.

Zhu, Z., Zhang, J., Wang, S., Xiao, Z., Huang, J., & Chen, H.-C. (2009). Involvement of left inferior frontal gyrus in sentence-level semantic integration. *NeuroImage*, *47*, 756–763.